PATENT
Customer No. 32425

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Peter DROGE, Nicole CHRIST and Elke LORBACH	Group Art Unit: 1636 Examiner: Q. Nguyen
Serial No.: 10/082,772	Atty. Dkt. No.: DEBE:008US/SLH
Filed: February 22, 2002	
For: SEQUENCE-SPECIFIC DNA RECOMBINATION IN EUKARYOTIC CELLS	Confirmation No.: 4391

CERTIFICATE OF ELECTRONIC TRANSMISSION
37 C.F.R. § 1.8

I hereby certify that this Appeal Brief is being electronically filed with the United States Patent and Trademark Office via EFS-Web on the date below:

October 6, 2008
Date

Steven L. Highlander

REPLY BRIEF

TABLE OF CONTENTS

Page

I. REAL PARTY IN INTEREST 1
II. RELATED APPEALS AND INTERFERENCES 1
III. STATUS OF THE CLAIMS 2
IV. STATUS OF THE AMENDMENTS 2
V. SUMMARY OF THE CLAIMED SUBJECT MATTER 2
VI. GROUND OF REJECTION TO BE REVIEWED ON APPEAL 2
VII. ARGUMENT 3
A. Standard of Review 3
B. Rejection of Claims 29 and 49-51 Under 35 U.S.C. §103 3
C. General Discussion of the References 4
D. Conclusion 8
VIII. APPENDIX A – APPEALED CLAIMS 9
IX. APPENDIX B – EVIDENCE CITED 12
X. APPENDIX C – RELATED PROCEEDINGS 13

PATENT
Customer No. 32425

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Peter DROGE, Nicole CHRIST and Elke LORBACH	Group Art Unit: 1636 Examiner: Q. Nguyen
Serial No.: 10/082,772	Atty. Dkt. No.: DEBE:008US/SLH
Filed: February 22, 2002	
For: SEQUENCE-SPECIFIC DNA RECOMBINATION IN EUKARYOTIC CELLS	Confirmation No.: 4391

REPLY BRIEF

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-01450

Dear Sir:

This Reply Brief is filed in response to the Examiner's Answer mailed on August 5, 2008. Appellant's brief is due October 6, 2008, October 5th being a Sunday. No fees are believed due in connection with this filing; however, should any fees be due, appellants authorize the Commissioner to debit Fulbright & Jaworski L.L.P. Deposit Account No. 50-1212/DEBE:008US.

I. Real Party In Interest

The real party in interest is the assignee, Peter Dröge.

II. Related Appeals and Interferences

There are no related appeals or interferences.

III. Status of the Claims

Claims 1-28 were filed with the original application, but these claims were canceled in a preliminary amendment in favor of new claims 29-60. Claims 29-51 and 58 were elected in a response to restriction requirement, and claims 31, 40-42, 52-57, 59 and 60 were canceled during prosecution. Thus, claims 29, 30, 32-39, 43-51 and 58 are pending in the application, stand rejected and are appealed. A copy of the appealed claims is attached as Appendix A.

IV. Status of the Amendments

No unentered amendments have been offered.

V. Summary of the Claimed Subject Matter

Independent claim 29, drawn to a method of sequence specific recombination of DNA in a eukaryotic cell using *attB, attP, attR* and *attL* sequences, along with wild-type or int-h or int-h/218 integrases, is supported in the application as filed at page 8, line 26, to page 9, line 6, and page 14, lines 9-14.

VI. Ground of Rejection to be Reviewed on Appeal

1. Are claims 29, 30, 32, 33, 36, 38, 44-48 and 58 obvious under 35 U.S.C. §103 over the combined disclosures of Crouzet *et al.* (Exhibit 1) and Christ & Dröge (Exhibit 2).

2. Are claims 29 and 43 obvious under 35 U.S.C. §103 over Crouzet *et al.*, Christ & Dröge and Capecchi *et al.* (Exhibit 3).

3. Are claims 29, 34-37 and 39 obvious under 35 U.S.C. §103 over Crouzet *et al.*, Christ & Dröge and Hartley *et al.* (Exhibit 4).

4. Are claims 29 and 49-51 obvious under 35 U.S.C. §103 over Crouzet *et al.*, Christ & Dröge, Hartley *et al.* and Calos *et al.* (Exhibit 7).

VII. Argument

A. Standard of Review

Findings of fact and conclusions of law by the U.S. Patent and Trademark Office must be made in accordance with the Administrative Procedure Act, 5 U.S.C. §706(A), (E), 1994. *Dickinson v. Zurko*, 527 U.S. 150, 158 (1999). Moreover, the Federal Circuit has held that findings of fact by the Board of Patent Appeals and Interferences must be supported by "substantial evidence" within the record. *In re Gartside*, 203 F.3d 1305, 1315 (Fed. Cir. 2000). In *In re Gartside*, the Federal Circuit stated that "the 'substantial evidence' standard asks whether a reasonable fact finder could have arrived at the agency's decision." *Id.* at 1312. Accordingly, it necessarily follows that an examiner's position on appeal must be supported by "substantial evidence" within the record in order to be upheld by the Board of Patent Appeals and Interferences.

B. Rejection of Claims 29 and 49-51 Under 35 U.S.C. §103

Claims 29 and 49-51 were newly rejected in the office action of February 21, 2008 over Crouzet, Calos, Hartley and Christ & Dröge. Inadvertantly, appellants did not identify that new and distinct rejection when filing their brief on appeal. Appellants submit, however, that this was in no way an intention to acquiesce to this rejection. Moreover, given the rejection of claim 29 in each of the other three grounds for rejection, this was in fact only a unique and new rejection of claims 49-51. Indeed, as stated, the rejection merely reapplies three references discussed elsewhere in the brief, with the addition of Calos. Yet this latter reference merely is cited for a teaching of modified integrases and an excision factor in the context of a ***second sequence specific recombination*** (see claims 49-51). Regardless, the teachings of Calos cannot correct the manifest defects already pointed out with respect to Crouzet, Hartley and Christ &

Dröge, and as such, the rejection relies on the same flawed premise as other rejections already traversed and argued many times over. Thus, for reasons given throughout the prosecution, in the previously filed appeal brief, and as set forth above, reversal of this rejection is requested as well.

C. General Discussion of the References

Crouzet *et al.* is directed, generally, to the provision of medicinal DNA molecules suitable for gene therapeutic use, more precisely, for *in vivo* gene transfer. For this purpose, molecules should lack any non-therapeutic regions, *e.g.*, origin of replication, resistance gene, non-relevant genes. Furthermore, these molecules should be prepared in high amounts and purity in supercoiled form appropriate for pharmaceutical use. To achieve these goals Crouzet *et al.* provide a method for the preparation and purification of mini-circles which fulfill the above mentioned criteria. Thus, the method according to their invention lies in the production of mini-circles by excision from a plasmid or from a chromosome by site-specific recombination whereby the site-specific recombination may be carried out *in vitro* in a host cell or *in vitro* on a plasmid preparation.

For a person of ordinary skill in the art, the most efficient way to produce plasmids in high amounts and purity is the production in bacteria. Bacteria can be easily grown to high cell densities in a short time. Furthermore, plasmids replicate easily in the bacterial host and can be maintained at high copy numbers in each cell. In addition, plasmids of high quality can be isolated in high yields by well known and established standard techniques. The ease and efficiency of production with high yield holds true for the generation of mini-circles by excision from a bacterial chromosome as well. The propagation of plasmids in bacteria and the excision of the mini-circles using the bacteriophage lambda are therefore the approaches described in the

examples provided by Crouzet *et al.* But, neither methods nor examples for the *in vivo* excision in eukaryotes is provided, and thus there was no proof given in the reference that lambda integrase would actually work in eukaryotes, much less whether one would have to provide accessory factors, and whether the yields and quality would be sufficient.

Moreover, it was completely unknown if supercoiled DNA could even be obtained following the teachinds of the reference. At that time, no one had shown functionality of lambda integrase in eukaryotes, so Crouzet *et al.* could not even rely on any published data. In eukaryotes, DNA in the nucleus is organized in form of chromatin. Nucleosomes, the primary repeating unit of chromatin, package DNA by wrapping it around an octamer of histone proteins. Crouzet *et al.* did not show that, after use of their cited standard DNA purification techniques, mini-circles in a supercoiled form would be obtained. Therefore, Crouzet *et al.*'s inclusion of any and all type of cell host, ***especially*** mammalian animal cells, was prophetic and completely unsupported from a scientific standpoint. The embodiment Crouzet *et al.* clearly had in mind was the production of mini-circles in bacteria – a notion further supported by the methods and descriptions given in the reference. For example, if excision from a plasmid should be performed in eukaryotes, the plasmid itself must comprise two origins of replication: one for propagation in bacteria to allow for the cloning of the plasmid, and one for replication in eukaryotes. However, the preferred or particular plasmid according to Crouzet *et al.*'s invention comprises ***only*** a bacterial origin of replication (col. 5, line 33; col. 6, line 14 and 48). In addition, for excision of the DNA molecule from the genome of the host cell, the descriptions relate ***solely*** to techniques used in prokaryotes (col. 8, lines 52–67, col. 9, lines 1–26). In col. 8, line 64, it even states "integrated in the genome of the bacterium." Thus, it cannot be argued that Crouzet *et al.* supports or suggests the use of eukaryotic cells.

Capecchi *et al.* merely provides better means to screen and select for successful homologous recombination events in eukaryotic cells by using positive-negative selection vectors. However, one has to differentiate between homologous and site-specific recombination. The former does not take place at specific site and relies on sequence similarities (or identity) of a piece of DNA introduced into a host cell and the DNA strand exchange can occur anywhere in the homologous regions. Also, it involves endogenous enzymes. On the other hand, site-specific recombination takes place at a specific site and makes use of enzymes (recombinase) which catalyse DNA strand exchange even between molecules that have only limited sequence homology. Capecchi *et al.* neither teach the use of site-specific recombination in eukaryotes, nor a combined approach of homologous recombination and site-specific recombination. Thus, the deficiencies set forth above with respect to Crouzet *et al.* remain.

Hartley *et al.* provides for recombination cloning of plasmids *in vitro*, and either the *in vitro* or *in vivo* selection of recombinatorial cloning products, thereby replacing more tedious, conventional cloning procedures which include digestion with restriction enzymes followed by ligation reactions. Hartley *et al.* teach the use of lambda integrase for site-specific recombination, but the actual recombinations were performed *in vitro* with purified wild-type lambda recombinase and purified accessory factors XIS and IHF. Also, Hartley *et al.* teach that selection for the actual recombinatorial cloning product (which is one of the product in the *in vitro* recombination mixture; others are by-product, co-integrate, insert and vector donor) can be performed either *in vitro* by using for example rare restriction enzyme site (col. 13, lines 12–26), or *in vivo* in host cells by using suitable selection markers (col. 9, lines 8–67, col. 10, lines 1–31). However, while hosts are defined as any prokaryotic or eukaryotic organism that can be a recipient of the recombinatorial cloning product (col. 8, lines 12–13), the actual recombination

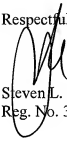
reaction is performed *in vitro*. Therefore, Hartley *et al.* do not teach site-specific recombinations in eukaryotic cells, nor do they teach means how to do so, nor do they even claim the execution of site-specific recombinations in eukaryotic cells for recombinatorial cloning.

Christ & Dröge teaches assays performed only in bacteria. Christ & Dröge do not teach site-specific recombinations in eukaryotic cells, nor do they provide means how to do so. In contrast to bacteria in which genomic or plasmid DNA is negatively supercoiled, DNA in the nucleus of eukaryotes is organized in form of chromatin. Nucleosomes, the primary repeating unit of chromatin package DNA by wrapping it around an octamer of histone proteins. Therefore, the DNA topology is quite different, and it was not obvious to an ordinary skilled artisan to deduce from the existing data that the mutant recombinases would work inside the nucleus with completely different DNA topology on an inter- or intramolecular level. It also was not clear that the mutant recombinases would even be transported into the nucleus, or that they would have any (or sufficient) biological activity or stability once in the nucleus, because the physiologicial conditions in a nucleus are different to the physiology in a bacterium or an *in vitro* system.

D. Conclusion

In light of the foregoing, appellant again respectfully submits that all pending claims are non-obvious under 35 U.S.C. §103. Therefore, it is respectfully requested that the Board reverse each of the pending rejections.

Respectfully submitted,



Date: October 6, 2008

Steven L. Highlander
Reg. No. 37,642

Fulbright & Jaworski L.L.P.
600 Congress Ave., Suite 2400
Austin TX 78701
512-474-5201

VIII. APPENDIX A – APPEALED CLAIMS

29. A method of sequence specific recombination of DNA in a eukaryotic cell, comprising:

(a) providing said eukaryotic cell, said cell comprising a first DNA segment integrated into the genome of said cell, said first DNA segment comprising an *att*B sequence according to SEQ ID NO:1 or a derivative thereof, an *att*P sequence according to SEQ ID NO:2 or a derivative thereof, an *att*L sequence according to SEQ ID NO:3 or a derivative thereof, or an *att*R sequence according to SEQ ID NO:4 or a derivative thereof;

(b) introducing a second DNA segment into said cell, wherein if said first DNA segment comprises an *att*B sequence according to SEQ ID NO:1 or a derivative thereof, said second DNA segment comprises an *att*P sequence according to SEQ ID NO:2 or a derivative thereof, wherein if said first DNA segment comprises an *att*P sequence according to SEQ ID NO:2 or a derivative thereof, said second DNA segment comprises an *att*B sequence according to SEQ ID NO:1 or a derivative thereof, wherein if said first DNA segment comprises an *att*L sequence according to SEQ ID NO:3 or a derivative thereof said second DNA segment comprises an *att*R sequence according to SEQ ID NO:4 or a derivative thereof, or wherein if said first DNA segment comprises an *att*R sequence according to SEQ ID NO:4 or a derivative thereof said second DNA segment comprises an *att*L sequence according to SEQ ID NO:3 or a derivative thereof; and

(c) further comprising providing to said cell a modified bacteriophage *lambda* integrase Int, wherein said modified Int is Int-h or Int-h/218, which induces sequence specific recombination through said *att*B and *att*P or *att*R and *att*L sequences.

30. The method of claim 29, wherein said first DNA segment was introduced into the genome of said cell by recombinant methods.

32. The method of claim 29, wherein said first DNA segment comprises an *att*B sequence

according to SEQ ID NO:1 or a derivative thereof, and said second DNA comprises an *att*P sequence according to SEQ ID NO:2 or a derivative thereof.

33. The method of claim 29, wherein said first DNA segment comprises an *att*P sequence according to SEQ ID NO:2 or a derivative thereof, and said second DNA comprises an *att*B sequence according to SEQ ID NO:1 or a derivative thereof.

34. The method of claim 29, wherein said first DNA segment comprises an *att*L sequence according to SEQ ID NO:3 or a derivative thereof, and said second DNA sequence comprises an *att*R sequence according to SEQ ID NO:4 or a derivative thereof, further comprising, in step (d), providing to said cell a Xis factor.

35. The method of claim 29, wherein said first DNA segment comprises an *att*R sequence according to SEQ ID NO:4 or a derivative thereof, and said second DNA sequence comprises an *att*L sequence according to SEQ ID NO:3 or a derivative thereof, further comprising, in step (d), providing to said cell a Xis factor.

36. The method of claim 29, further comprising providing to said cell a third DNA segment comprising an Int gene.

37. The method of claim 36, further comprising providing to said cell a fourth DNA segment comprising a Xis factor gene, respectively.

38. The method of claim 36, wherein said third DNA segment further comprises a regulatory sequence effecting a spatial and/or temporal expression of the Int gene.

39. The method of claim 37, wherein said fourth DNA segment further comprises a regulatory sequence effecting a spatial and/or temporal expression of the Xis factor gene.

43. The method according to claim 29, wherein said first and/or second DNA segment further comprise a sequence effecting integration of said first and/or second DNA segment into the genome of said cell by homologous recombination.

44. The method of claim 29, wherein said first and/or second DNA segment further

comprises a sequence coding for a polypeptide of interest.

45. The method of claim 44, wherein said polypeptide of interest is a structural protein, an endogenous or exogenous enzyme, a regulatory protein or a marker protein.

46. The method of claim 29, wherein said first and second DNA segment are introduced into the eukaryotic cell on the same DNA molecule.

47. The method of claim 29, wherein said eukaryotic cell is a mammalian cell.

48. The method of claim 47, wherein said mammalian cell is a human, simian, mouse, rat, rabbit, hamster, goat, bovine, sheep or pig cell.

49. The method of claim 29, further comprising:

(d) performing a second sequence specific recombination of DNA by Int-h or Int-h/218 and a Xis factor after the steps (a)-(c), wherein said first DNA sequence comprises said *att*B sequence according to SEQ ID NO:1 or a derivative thereof and said second DNA sequence comprises the *att*P sequence according to SEQ ID NO:2 or a derivative thereof, or wherein said first DNA sequence comprises said *att*P sequence according to SEQ ID NO:2 or a derivative thereof and said second DNA sequence comprises the *att*B sequence according to SEQ ID NO:1 or a derivative thereof .

50. The method of claim 49, further introducing a further DNA sequence into said cells, the further DNA sequence comprising a Xis factor gene.

51. The method of claim 50, wherein said further DNA sequence comprises further a regulatory DNA sequence effecting a spatial and/or temporal expression of said Xis factor gene.

58. An isolated eukaryotic cell obtainable according to the method of claim 29.

IX. APPENDIX B – EVIDENCE CITED

Exhibit 1 – Crouzet *et al.*

Exhibit 2 – Christ & Dröge *et al.*

Exhibit 3 – Capecchi *et al.*

Exhibit 4 – Hartley *et al.*

Exhibit 5 – Lange-Gustafson *et al.*

Exhibit 6 – Declaration of Peter Dröge

Exhibit 7 – Calos *et al.*

X. APPENDIX C – RELATED PROCEEDINGS

None

EXHIBIT 1

US 6143530A

United States Patent [19]

Crouzet et al.

[11] Patent Number: 6,143,530

[45] Date of Patent: Nov. 7, 2000

[54] **CIRCULAR DNA EXPRESSION CASSETTES FOR IN VIVO GENE TRANSFER**

[75] Inventors: **Joël Crouzet**, Sceaux; **Daniel Scherman**, Paris; **Béatrice Cameron**, Paris; **Pierre Wils**, Paris; **Anne-Marie Darquet**, Vitry sur Seine, all of France

[73] Assignee: **Rhone-Poulenc Rorer S.A.**, Antony, France

[21] Appl. No.: **08/894,511**

[22] PCT Filed: **Feb. 21, 1996**

[86] PCT No.: **PCT/FR96/00274**

§ 371 Date: **Aug. 19, 1997**

§ 102(e) Date: **Aug. 19, 1997**

[87] PCT Pub. No.: **WO96/26270**

PCT Pub. Date: **Aug. 29, 1996**

[30] **Foreign Application Priority Data**

Feb. 23, 1995 [FR] France 95 02117

[51] **Int. Cl.**[7] **C12N 15/64**; C12N 15/79; C12N 15/70; C12N 1/21; C12N 5/10

[52] **U.S. Cl.** **435/91.42**; 435/91.1; 435/91.4; 435/320.1; 435/325; 435/252.3; 435/252.33; 435/254.11; 435/455; 514/44

[58] **Field of Search** 514/44; 435/320.1, 435/325, 252.3, 252.33, 254.11, 91.1, 91.4, 91.42, 455

[56] **References Cited**

U.S. PATENT DOCUMENTS

5,227,288	7/1993	Blattner	435/6
5,401,632	3/1995	Wang et al.	435/6

FOREIGN PATENT DOCUMENTS

300 422	1/1989	European Pat. Off. .
350 341	1/1990	European Pat. Off. .
WO 94/09127	4/1994	WIPO .
WO 96/05297	2/1996	WIPO .

OTHER PUBLICATIONS

Takabatake et al., The use of purine–rich oligonucleotides in triplex–mediated DNA isolation and generation of unidirectional deletions, Nucletic Acids Research, 20(21), 5853–5854 (1992).

Ito et al., Sequence–specific DNA purification by triplex affinity capture, Proc. Natl. Acad. Sci. USA, 89, 495–498 (1992).

Mizuuchi et al., The extent of DNA sequence required for a functional bacterial attachment site of phage lambda, Nucleic–Acids Research, 13(4) 1193–1208 (1985).

Hasan et al., Control of cloned gene expression by promoter inversion in vivo: construction of improved vectors with a multiple cloning site and the ptac promoter, Gene, 56, 145–151 (1987).

Su et al., Selective Binding of *Escherichia coli* RNA Polymerase to Topoisomers of Minicircles Carrying the TAC16 and TAC17 Promoters, Journal of Biological Chemistry 269(18), 13511–13521 (1994).

Backman et al., Use of Synchronous Site–Specific Recombination In Vivo To Regulate Gene Expression, Biotechnology, 2(12) 1045–1049 (1984).

Eberl et al., Analysis of the multimer resolution system encoded by the parCBA operon of broad–host–range plasmid RP4, Molecular Microbiology 12(1), 131–141 (1994).

Stark et al., Catalysis by site–specific recombinases, Trenda in Genetics, 8(12), 432–439 (1992).

Orkin et al., "Report and recommendations of the panel to assess the NIH investment in research on gene therapy", issued by the U.S. National Institutes of Health, Dec. 1995.

Ben–Yedidia et al., "Design of peptide and polypeptide vaccines", Curr. Opin. Biotechnol. 8: 442–448, Aug. 1997.

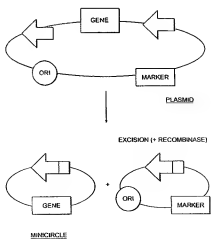
Primary Examiner—Scott D. Priebe

Attorney, Agent, or Firm—Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.

[57] **ABSTRACT**

Double-stranded DNA molecules characterised in that they are circular and in that they essentially include one or more genes of interest.

43 Claims, 11 Drawing Sheets



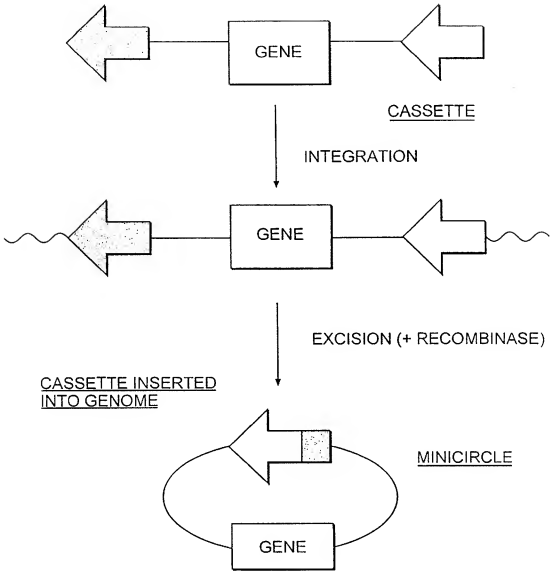

GENE
CASSETTE
INTEGRATION
GENE
EXCISION (+ RECOMBINASE)
CASSETTE INSERTED INTO GENOME
MINICIRCLE
GENE

FIG. 1

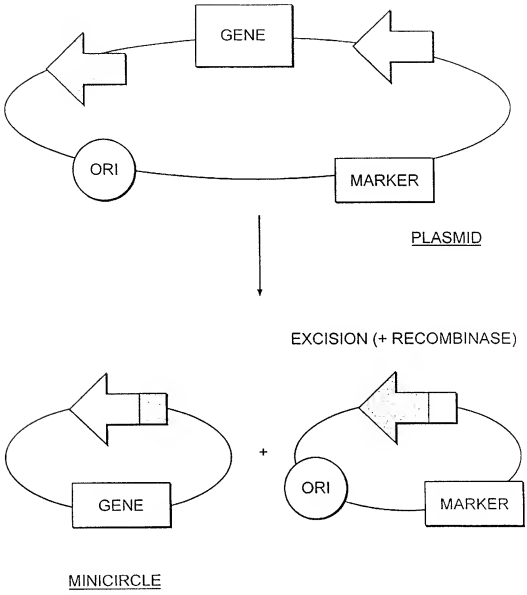

GENE
ORI
MARKER
PLASMID
EXCISION (+ RECOMBINASE)
GENE
+
ORI
MARKER
MINICIRCLE

FIG. 2

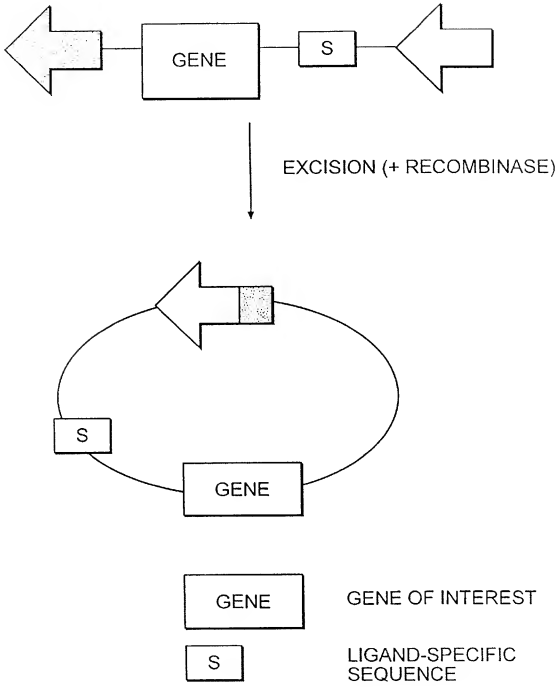

GENE
S
EXCISION (+ RECOMBINASE)
S
GENE
GENE
GENE OF INTEREST
S
LIGAND-SPECIFIC SEQUENCE

FIG. 3

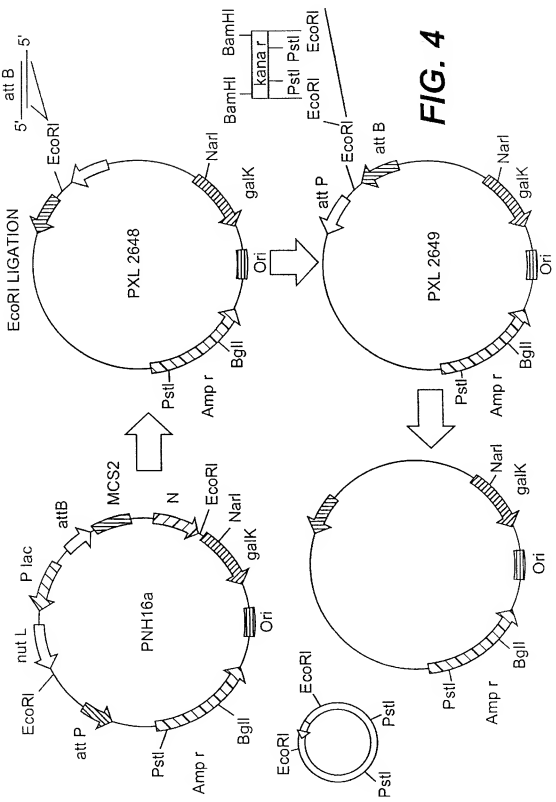
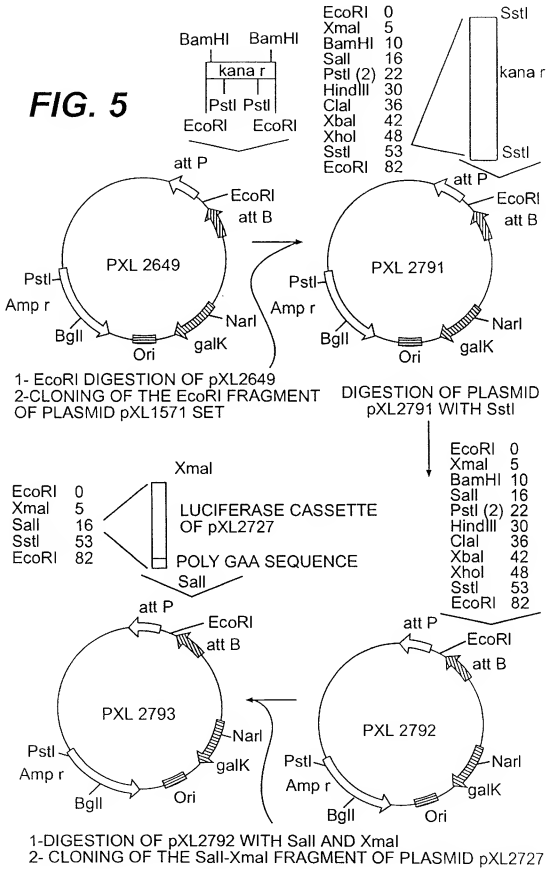

EcoRI LIGATION
PXL 2648
PXL 2649
PNH16a
MCS2
attB
P lac
nut L
att P
att B
EcoRI
NarI
galK
Ori
BglI
PstI
Amp r
N
BamHI
kana r
5'
3'

FIG. 4

FIG. 5


BamHI
BamHI
kana r
PstI
PstI
EcoRI
EcoRI
EcoRI 0
Xmal 5
BamHI 10
Sall 16
PstI (2) 22
HindIII 30
ClaI 36
XbaI 42
XhoI 48
SstI 53
EcoRI 82
SstI
kana r
SstI
att P
EcoRI
att B
PXL 2649
PstI
Amp r
BglI
Ori
NarI
galK
PXL 2791
1- EcoRI DIGESTION OF pXL2649
2-CLONING OF THE EcoRI FRAGMENT
OF PLASMID pXL1571 SET
DIGESTION OF PLASMID
pXL2791 WITH SstI
Xmal
EcoRI 0
Xmal 5
Sall 16
SstI 53
EcoRI 82
LUCIFERASE CASSETTE
OF pXL2727
POLY GAA SEQUENCE
Sall
PXL 2793
PXL 2792
1-DIGESTION OF pXL2792 WITH SalI AND XmaI
2- CLONING OF THE SalI-XmaI FRAGMENT OF PLASMID pXL2727

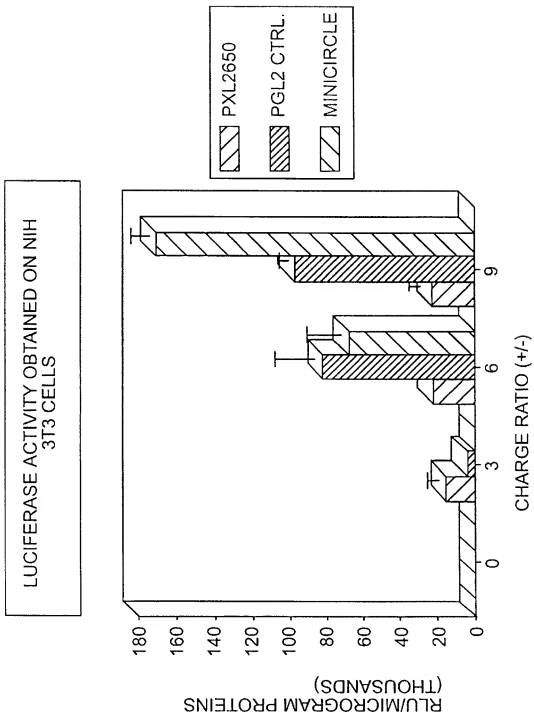

LUCIFERASE ACTIVITY OBTAINED ON NIH 3T3 CELLS
PXL2650
PGL2 CTRL.
MINICIRCLE
RLU/MICROGRAM PROTEINS (THOUSANDS)
0
20
40
60
80
100
120
140
160
180
0
3
6
9
CHARGE RATIO (+/-)

FIG. 6

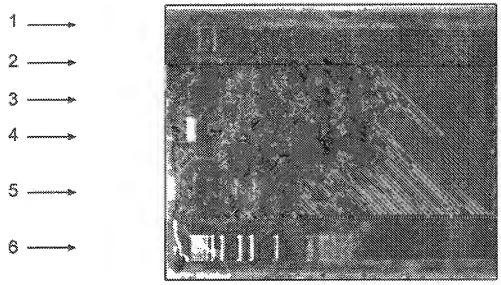


FIG. 7

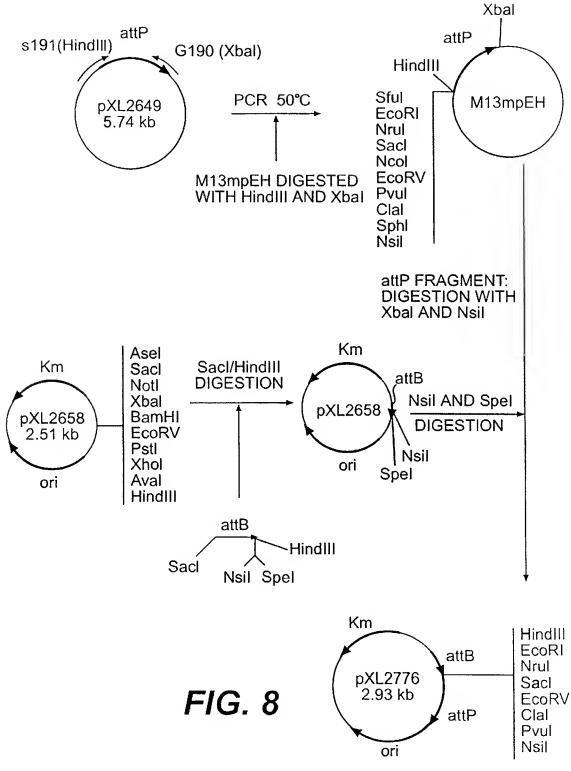

attP
s191(HindIII)
G190 (XbaI)
pXL2649
5.74 kb
PCR 50°C
M13mpEH DIGESTED
WITH HindIII AND XbaI
XbaI
attP
HindIII
M13mpEH
SfuI
EcoRI
NruI
SacI
NcoI
EcoRV
PvuI
ClaI
SphI
NsiI
attP FRAGMENT:
DIGESTION WITH
XbaI AND NsiI
Km
pXL2658
2.51 kb
ori
AseI
SacI
NotI
XbaI
BamHI
EcoRV
PstI
XhoI
AvaI
HindIII
SacI/HindIII
DIGESTION
Km
attB
pXL2658
ori
NsiI AND SpeI
DIGESTION
NsiI
SpeI
attB
HindIII
SacI
NsiI SpeI
Km
pXL2776
2.93 kb
attB
attP
ori
HindIII
EcoRI
NruI
SacI
EcoRV
ClaI
PvuI
NsiI

FIG. 8

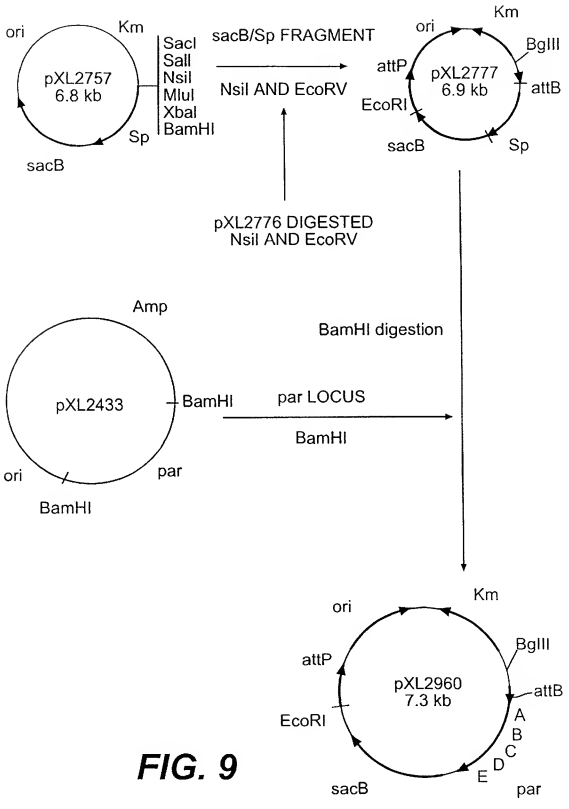

ori
Km
pXL2757
6.8 kb
SacI
SalI
NsiI
MluI
XbaI
BamHI
Sp
sacB
sacB/Sp FRAGMENT
NsiI AND EcoRV
ori
Km
BglII
attP
pXL2777
6.9 kb
attB
EcoRI
sacB
Sp
pXL2776 DIGESTED
NsiI AND EcoRV
BamHI digestion
Amp
pXL2433
BamHI
par LOCUS
BamHI
ori
par
BamHI
ori
Km
BglII
attP
pXL2960
7.3 kb
attB
EcoRI
A
B
C
D
E
sacB
par

FIG. 9

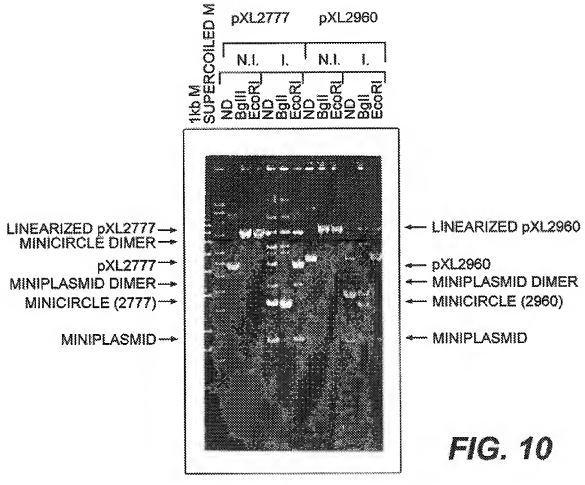

1kb M
SUPERCOILED M
pXL2777
pXL2960
N.I.
I.
N.I.
I.
ND
BglII
EcoRI
ND
BglII
EcoRI
ND
BglII
EcoRI
ND
BglII
EcoRI
LINEARIZED pXL2777
MINICIRCLE DIMER
pXL2777
MINIPLASMID DIMER
MINICIRCLE (2777)
MINIPLASMID
LINEARIZED pXL2960
pXL2960
MINIPLASMID DIMER
MINICIRCLE (2960)
MINIPLASMID

FIG. 10

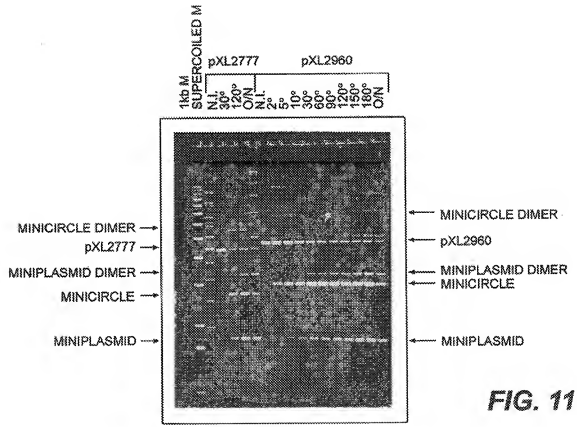

1kb M
SUPERCOILED M
pXL2777
N.I.
30°
120°
O/N
pXL2960
N.I.
2°
5°
10°
30°
60°
90°
120°
150°
180°
O/N
MINICIRCLE DIMER
pXL2777
MINIPLASMID DIMER
MINICIRCLE
MINIPLASMID
MINICIRCLE DIMER
pXL2960
MINIPLASMID DIMER
MINICIRCLE
MINIPLASMID

FIG. 11

CIRCULAR DNA EXPRESSION CASSETTES FOR IN VIVO GENE TRANSFER

Gene therapy consists in correcting a deficiency or an abnormality by introducing genetic information into the affected cell or organ. This information may be introduced either in vitro into a cell extracted from the organ and then reinjected into the body, or in vivo, directly into the tissue concerned. Being a high molecular weight, negatively charged molecule, DNA has difficulties in passing spontaneously through the phospholipid cell membranes. Different vectors are hence used in order to permit gene transfer: viral vectors on the one hand, natural or synthetic, chemical and/or biochemical vectors on the other hand. Viral vectors (retroviruses, adenoviruses, adeno-associated viruses, etc.) are very effective, in particular in passing through membranes, but present a number of risks, such as pathogenicity, recombination, replication, immunogenicity, etc. Chemical and/or biochemical vectors enable these risks to be avoided (for reviews, see Behr, 1993, Cotten and Wagner, 1993). These vectors are, for example, cations (calcium-phosphate, DEAE-dextran, etc.) which act by forming precipitates with DNA, which precipitates can be "phagocytosed" by the cells. They can also be liposomes in which DNA is incorporated and which fuse with the plasma membrane. Synthetic gene transfer vectors are generally lipids or cationic polymers which complex DNA and form a particle therewith carrying positive surface charges. These particles are capable of interacting with the negative charges of the cell membrane and then of crossing the latter. Dioctadecylamidoglycylspermine (DOGS, Transfectam™) or N-[1-(2,3-dioleyloxy)propyl]-N,N,N-trimethylammonium chloride (DOTMA, Lipofectin™) may be mentioned as examples of such vectors. Chimeric proteins have also been developed: they consist of a polycationic portion which condenses DNA, linked to a ligand which binds to a membrane receptor and carries the complex into the cells by endocytosis. It is thus theoretically possible to "target" a tissue or certain cell populations so as to improve the in vivo bioavailability of the transferred gene.

However, the use of chemical and/or biochemical vectors or of naked DNA implies the possibility of producing large amounts of DNA of pharmacological purity. In effect, in these gene therapy techniques, the medicinal product consists of the DNA itself, and it is essential to be able to manufacture, in appropriate amounts, DNAs having suitable properties for therapeutic use in man.

The plasmids currently used in gene therapy carry (i) an origin of replication, (ii) a marker gene such as a gene for resistance to an antibiotic (kanamycin, ampicillin, etc.) and (iii) one or more transgenes with sequences required for their expression (enhancer(s), promoter(s), polyadenylation sequences, etc.). These plasmids currently used in gene therapy (in clinical trials such as the treatment of melanomas, Nabel et al., 1992, or in experimental studies) display, however, some drawbacks associated, in particular, with their dissemination in the body. Thus, as a result of this dissemination, a competent bacterium present in the body can, at a low frequency, receive this plasmid. The chance of this occurring is all the greater for the fact that the treatment in question entails in vivo gene therapy in which the DNA may be disseminated in the patient's body and may come into contact with bacteria which infect this patient or alternatively with bacteria of the commensal flora. If the bacterium which is a recipient of the plasmid is an enterobacterium such as *E. coli*, this plasmid may replicate. Such an event then leads to the dissemination of the therapeutic gene.

Inasmuch as the therapeutic genes used in gene therapy treatments can code, for example, for a lymphokine, a growth factor, an anti-oncogene, or a protein whose function is lacking in the host and hence enables a genetic defect to be corrected, the dissemination of some of these genes could have unforeseeable and worrying effects (for example if a pathogenic bacterium were to acquire the gene for a human growth factor). Furthermore, the plasmids used in non-viral gene therapy also possess a marker for resistance to an antibiotic (ampicillin, kanamycin, etc.). Hence the bacterium acquiring such a plasmid has an undeniable selective advantage, since any therapeutic antibiotic treatment using an antibiotic of the same family as the one selecting the resistance gene of the plasmid will lead to the selection of the plasmid in question. In this connection, ampicillin belongs to the β-lactams, which is the family of antibiotics most widely used in the world. It is hence necessary to seek to limit as far as possible the dissemination of the therapeutic genes and the resistance genes. Moreover, the genes carried by the plasmid, corresponding to the vector portion of the plasmid (function(s) required for replication, resistance gene), also run the risk of being expressed in the transfected cells. There is, in effect, a transcription background, which cannot be ruled out, due to the host's expression signals on the plasmid. This expression of exogenous proteins may be thoroughly detrimental in a number of gene therapy treatments, as a result of their potential immunogenicity and hence of the attack of the transfected cells by the immune system.

Hence it is especially important to be able to have at one's disposal medicinal DNA molecules having a genetic purity suitable for therapeutic use. It is also especially important to have at one's disposal methods enabling these DNA molecules to be prepared in amounts appropriate for pharmaceutical use. The present invention provides a solution to these problems.

The present invention describes, in effect, DNA molecules which can be used in gene therapy, having greatly improved genetic purity and impressive properties of bioavailability. The invention also describes an especially effective method for the preparation of these molecules and for their purification.

The present invention lies, in particular, in the development of DNA molecules which can be used in gene therapy, virtually lacking any non-therapeutic region. The DNA molecules according to the invention, also designated minicircles on account of their circular structure, their small size and their supercoiled form, display many advantages.

They make it possible, in the first place, to eliminate the risks associated with dissemination of the plasmid, such as (1) replication and dissemination which may lead to an uncontrolled overexpression of the therapeutic gene, (2) the dissemination and expression of resistance genes, and (3) the expression of genes present in the non-therapeutic portion of the plasmid, which are potentially immunogenic and/or inflammatory, and the like. The genetic information contained in the DNA molecules according to the invention is limited, in effect, essentially to the therapeutic gene(s) and to the signals for regulation of its/their expression (neither origin of replication, nor gene for resistance to an antibiotic, and the like). The probability of these molecules (and hence of the genetic information they contain) being transferred to a microorganism and being stably maintained is almost zero.

Furthermore, due to their small size, DNA molecules according to the invention potentially have better bioavailability in vivo. In particular, they display improved capacities for cell penetration and cellular distribution. Thus, it is

recognized that the coefficient of diffusion in the tissues is inversely proportional to the molecular weight (Jain, 1987). Similarly, at cellular level, high molecular weight molecules have inferior permeability through the plasma membrane. In addition, for the plasmid to progress to the nucleus, which is essential for its expression, high molecular weight is also a drawback, the nuclear pores imposing a size limit for diffusion to the nucleus (Landford et al., 1986). The elimination of the non-therapeutic portions of the plasmid (origin of replication and resistance gene in particular) according to the invention also enables the size of the DNA molecules to be decreased. This decrease may be estimated at a factor of 2, reckoning, for example, 3 kb for the origin of replication and the resistance marker (vector portion) and 3 kb for the transgene with the sequences required for its expression. This decrease (i) in molecular weight and (ii) in negative charge endows the molecules of the invention with improved capacities for tissue, cellular and nuclear diffusion and bioavailability.

Hence a first subject of the invention lies in a double-stranded DNA molecule having the following features: it is circular in shape and essentially comprises one or more genes of interest. As stated above, the molecules of the invention essentially lack non-therapeutic regions, and especially an origin of replication and/or a marker gene. In addition, they are advantageously in supercoiled form.

The present invention is also the outcome for the development of a method, of constructions and of cell hosts which are specific and especially effective for the production of these therapeutic DNA molecules. More especially, the method according to the invention lies in the production of therapeutic DNA molecules defined above, by excision from a plasmid or from a chromosome by site-specific recombination. The method according to the invention is especially advantageous, since it does not necessitate a prior step of purification of the plasmid, is very specific, especially effective, does not decrease the amounts of DNA produced and leads directly to therapeutic molecules of very great genetic purity and of great bioavailability. This method leads, in effect, to the generation of circular DNA molecules (minicircles) essentially containing the gene of interest and the regulator sequences permitting its expression in the cells, tissue, organ or apparatus, or even the whole body, in which the expression is desired. In addition, these molecules may then be purified by standard techniques.

The site-specific recombination may be carried out by means of various systems which lead to site-specific recombination between sequences. More preferably, the site-specific recombination in the method of the invention is obtained by means of two specific sequences which are capable of recombining with one another in the presence of a specific protein, generally designated recombinase. For this reason, the DNA molecules according to the invention generally comprise, in addition, a sequence resulting from this site-specific recombination. The sequences permitting the recombination used in the context of the invention generally comprise from 5 to 100 base pairs, and more preferably fewer than 50 base pairs.

The site-specific recombination may be carried out in vivo (that is to say in the host cell) or in vitro (that is to say on a plasmid preparation).

In this connection, the present invention also provides particular genetic constructions suitable for the production of the therapeutic DNA molecules defined above. These genetic constructions, or recombinant DNAs, according to the invention comprise, in particular, the gene or genes of interest flanked by the two sequences permitting site-specific recombination, positioned in the direct orientation. The position in the direct orientation indicates that the two sequences follow the same 5'-3' polarity in the recombinant DNA according to the invention. The genetic constructions of the invention can be double-stranded DNA fragments (cassettes) essentially composed of the elements mentioned above. These cassettes can be used for the construction of cell hosts having these elements integrated in their genome (FIG. 1). The genetic constructions of the invention can also be plasmids, that is to say any linear or circular DNA molecule capable of replicating in a given host cell, containing the gene or genes of interest flanked by the two sequences permitting site-specific recombination, positioned in the direct orientation. The construction can be, more specifically, a vector (such as a cloning and/or expression vector), a phage, a virus, and the like. These plasmids of the invention may be used to transform any competent cell host for the purpose of the production of minicircles by replication of the plasmid followed by excision of the minicircle (FIG. 2).

In this connection, another subject of the invention lies in a recombinant DNA comprising one or more genes of interest, flanked by two sequences permitting site-specific recombination, positioned in the direct orientation.

The recombinant DNA according to the invention is preferably a plasmid comprising at least:

a) an origin of replication and optionally a marker gene,

b) two sequences permitting a site-specific recombination, positioned in the direct orientation, and,

c) placed between said sequences b), one or more genes of interest.

The specific recombination system present in the genetic constructions according to the invention can be of different origins. In particular, the specific sequences and the recombinases used can belong to different structural classes, and in particular to the integrase family of bacteriophage λ or to the resolvase family of the transposon Tn3.

Among recombinases belonging to the integrase family of bacteriophage λ, there may be mentioned, in particular, the integrase of the phages lambda (Landy et al., Science 197 (1977) 1147), P22 and Φ80 (Leong et al., J. Biol. Chem. 260 (1985) 4468), HP1 of *Haemophilus influenza* (Hauser et al., J. Biol. Chem. 267 (1992) 6859), the Cre integrase of phage P1, the integrase of the plasmid pSAM2 (EP 350,341) or alternatively the FLP recombinase of the 2μ plasmid. When the DNA molecules according to the invention are prepared by recombination by means of a site-specific system of the integrase family of bacteriophage lambda, the DNA molecules according to the invention generally comprise, in addition, a sequence resulting from the recombination between two att attachment sequences of the corresponding bacteriophage or plasmid.

Among recombinases belonging to the family of the transposon Tn3, there may be mentioned, in particular, the resolvase of the transposon Tn3 or of the transposons Tn21 and Tn522 (Stark et al., 1992); the Gin invertase of bacteriophage mu or alternatively the resolvase of plasmids, such as that of the par fragment of RP4 (Albert et al., Mol. Microbiol. 12 (1994) 131). When the DNA molecules according to the invention are prepared by recombination by means of a site-specific system of the family of the transposon Tn3, the DNA molecules according to the invention generally comprise, in addition, a sequence resulting from the recombination between two recognition sequences of the resolvase of the transposon in question.

According to a particular embodiment, in the genetic constructions of the present invention, the sequences per-

mitting site-specific recombination are derived from a bacteriophage. More preferably, these latter are attachment sequences (attP and attB sequences) of a bacteriophage, or derived sequences. These sequences are capable of recombining specifically with one another in the presence of a recombinase designated integrase. The term derived sequence includes the sequences obtained by modification(s) of the attachment sequences of the bacteriophages, which retain the capacity to recombine specifically in the presence of the appropriate recombinase. Thus, such sequences can be reduced fragments of these sequences or, on the contrary, fragments extended by the addition of other sequences (restriction sites, and the like). They can also be variants obtained by mutation(s), in particular by point mutation(s). The terms attP and attB sequences of a bacteriophage or of a plasmid denote, according to the invention, the sequences of the recombination system specific to said bacteriophage or plasmid, that is to say the attP sequence present in said phage or plasmid and the corresponding chromosomal attB sequence.

By way of preferred examples, there may be mentioned, in particular, the attachment sequences of the phages lambda, P22, Φ80, P1 and HP1 of *Haemophilus influenzae* or alternatively of plasmid pSAM2 or the 2μ plasmid. These sequences are advantageously chosen from all or part of the sequences SEQ ID No. 1, SEQ ID No. 2, SEQ ID No. 6, SEQ ID No. 7, SEQ ID No. 8, SEQ ID No. 9, SEQ ID No. 10, SEQ ID No. 11, SEQ ID No. 12, SEQ ID No. 13 and SEQ ID No. 14. These sequences comprise, in particular, the central region homologous to the attachment sequences of these phages.

In this connection, a preferred plasmid according to the present invention comprises

(a) a bacterial origin of replication and optionally a marker gene,

(b) the attP and attB sequences of a bacteriophage selected from the phages lambda, P22, Φ80, HP1 and P1 or of plasmid pSAM2 or the 2μ plasmid, or derived sequences; and,

(c) placed between said sequences b), one or more genes of interest.

According to an especially preferred embodiment, the sequences in question are the attachment sequences (attP and attB) of phage lambda. Plasmids carrying these sequences are, in particular, the plasmids pXL2648, pXL2649 or pXL2650. When these plasmids are brought, in vivo or in vitro, into contact with the integrase of phage lambda, the sequences recombine with one another to generate in vivo or in vitro, by excision, a minicircle according to the invention essentially comprising the elements (c), that is to say the therapeutic portion (FIG. 2).

Still according to a particular embodiment of the invention, the sequences permitting site-specific recombination are derived from the loxP region of phage P1. This region is composed essentially of two inverted repeat sequences capable of recombining specifically with one another in the presence of a protein, designated Cre (Sternberg et al., J. Mol. Biol. 150 (1971) 467). In a particular variant, the invention hence relates to a plasmid comprising (a) a bacterial origin of replication and optionally a marker gene; (b) the inverted repeat sequences of bacteriophage P1 (loxP region); and (c), placed between said sequences (b), one or more genes of interest.

According to another particular embodiment, in the genetic constructions of the present invention, the sequences permitting site-specific recombination are derived from a transposon. More preferably, the sequences in question are recognition sequences of the resolvase of a transposon, or derived sequences. By way of preferred examples, there may be mentioned, in particular, the recognition sequences of the transposons Tn3, Tn21 and Tn522. By way of a preferred example, there may be mentioned the sequence SEQ ID No. 15 or a derivative of the latter (see also Sherrat, P. 163–184, Mobile DNA, Ed. D. Berg and M. Howe, American Society for Microbiology, Washington D.C. 1989).

According to another especially advantageous variant, the plasmids of the invention comprise, in addition, a multimer resolution sequence. This is preferably the mrs (multimer resolution system) sequence of the plasmid RK2. More preferably, the invention relates to a plasmid comprising:

(a) a bacterial origin of replication and optionally a marker gene,

(b) the attP and attB sequences of a bacteriophage, in the direct orientation, selected from the phages lambda, P22, Φ80, HP1 and P1 or of plasmid pSAM2 or the 2μ plasmid, or derived sequences; and,

(c) placed between said sequences b), one or more genes of interest and the mrs sequence of plasmid RK2.

This embodiment is especially advantageous. Thus, when plasmids pXL2649 or pXL2650 are brought into contact with the integrase of the bacteriophage in vivo, the sequences recombine to generate the minicircle and the miniplasmid, but also multimeric or topological forms of minicircle or of miniplasmid. It is especially advantageous to be able to decrease the concentration of these forms in order to increase the production and facilitate the purification of minicircle.

The multimeric forms of plasmids are known to a person skilled in the art. For example, the cer fragment of ColE1 (Summers et al., 1984 Cell 36 p. 1097) or the mrs site of the par locus of RK2 (L. Ebert 1994 Mol. Microbiol. 2 p. 131) permit the resolution of multimers of plasmids and participate in an enhanced stability of the plasmid. However, whereas resolution at the cer site requires four proteins encoded by the *E. coli* genome (Colloms et al., 1990 J. Bacteriol. 172 p. 6973), resolution at the mrs site requires only the ParA protein for which the parA gene is mapped on the par locus of RK2. As a result, it would appear advantageous to use all or a portion of the par locus containing parA and the mrs sequence. For example, the mrs sequence may be placed between the attB and attP sequences of phage lambda, and the parA gene be expressed in trans or in cis from its own promoter or from an inducible promoter. In this connection, a particular plasmid of the invention comprises:

(a) a bacterial origin of replication and optionally a marker gene,

(b) the attP and attB sequences of a bacteriophage, in the direct orientation, selected from the phages lambda, P22, Φ80, HP1 and P1 or of plasmid pSAM2 or the 2μ plasmid, or derived sequences,

(c) placed between said sequences b), one or more genes of interest and the mrs sequence of plasmid RK2, and

(d) the parA gene of plasmid RK2.

One such plasmid is, in particular, the plasmid pXL2960 described in the examples. It may be employed, and can enable minicircle to be produced exclusively in monomeric form.

According to another advantageous variant, the plasmids of the invention comprise two sets of site-specific recombination sequences from a different family. These advantageously comprise a first set of integrase-dependent sequences and a second set of parA-dependent sequences. The use of two sets of sequences enables the production

yields of minicircles to be increased when the first site-specific recombination is incomplete. Thus, when plasmids pXL2650 or pXL2960 are brought into contact with the integrase of the bacteriophage in vivo, the sequences recombine to generate the miniplasmid and the minicircle, but this reaction is not complete (5 to 10% of initial plasmid may be left). The introduction, in proximity to each of the att sequences of phage lambda, of an mrs sequence of RK2 enables the production of minicircles to be increased. Thus, after induction of the integrase of phage lambda and Int-dependent recombination, the unrecombined molecules will be able to come under the control of the ParA protein of RK2 and to recombine at the mrs sites. Conversely, after induction of the ParA protein and ParA-dependent recombination, the unrecombined molecules will be able to come under the control of the integrase of phage lambda and will be able to recombine at the att sites. Such constructions thus make it possible to produce minicircle and negligible amounts of unrecombined molecules. The att sequences, like the mrs sequences, are in the direct orientation, and the int and parA genes may be induced simultaneously or successively from the same inducible promoter or from two inducible promoters. Preferably, the sequences in question are the attB and attP attachment sequences of phage lambda in the direct orientation and two mrs sequences of RK2 in the direct orientation.

As stated above, another aspect of the present invention lies in a method for the production of therapeutic DNA molecules defined above, by excision, from a plasmid or chromosome, by site-specific recombination.

Another subject of the present invention hence lies in a method for the production of a DNA molecule (minicircle) as defined above, according to which a culture of host cells containing a recombinant DNA as defined above is brought into contact with the recombinase enabling site-specific recombination to be induced. More preferably, the culture and recombinase are brought into contact either by transfection or infection with a plasmid or a phage containing the gene for said recombinase; or by induction of the expression of a gene coding for said recombinase, present in the host cell. As mentioned below, this gene may be present in the host cell in integrated form in the genome, on a replicative plasmid or alternatively on the plasmid of the invention, in the non-therapeutic portion.

To permit the production of the minicircles according to the invention by site-specific recombination in vivo, the recombinase used must be introduced into, or induced in, cells or the culture medium at a particular instant. For this purpose, different methods may be used. According to a first method, a host cell is used containing the recombinase gene in a form permitting its regulated expression. It may, in particular, be introduced under the control of a promoter or of a system of inducible promoters, or alternatively in a temperature-sensitive system. In particular, the gene may be present in a temperature-sensitive phage, latent during the growth phase, and induced at a suitable temperature (for example lysogenic phage lambda Xis^{-} cI857). The cassette for expression of the recombinase gene may be carried by a plasmid, a phage or even by the plasmid of the invention, in the non-therapeutic region. It may be integrated in the genome of the host cell or maintained in replicative form. According to another method, the cassette for expression of the gene is carried by a plasmid or a phage used to transfect or infect the cell culture after the growth phase. In this case, it is not necessary for the gene to be in a form permitting its regulated expression. In particular, any constitutive promoter may be used. The cell may also be brought into contact with the recombinase in vitro, on a plasmid preparation, by direct incubation with the protein.

It is preferable, in the context of the present invention, to use a host cell capable of expressing the recombinase gene in a regulated manner. This embodiment, in which the recombinase is supplied directly by the host cell after induction, is especially advantageous. In effect, it suffices simply to place the cells in culture at the desired time under the conditions for expression of the recombinase gene (permissive temperature for a temperature-sensitive gene, addition of an inducer for a regulable promoter, and the like) in order to induce the site-specific recombination in vivo and thus the excision of the minicircle of the invention. In addition, this excision takes place in especially high yields, since all the cells in culture express the recombinase, which is not necessarily the case if a transfection or an infection has to be carried out in order to transfer the recombinase gene.

According to a first embodiment, the method of the invention comprises the excision of the molecules of therapeutic DNA by site-specific recombination from a plasmid. This embodiment employs the plasmids described above permitting, in a first stage, replication in a chosen host, and then, in a second stage, the excision of the non-therapeutic portions of said plasmid (in particular the origin of replication and the resistance gene) by site-specific recombination, generating the circular DNA molecules of the invention. To carry out the method, different types of plasmid may be used, and especially a vector, a phage or a virus. A replicative vector is preferably used.

Advantageously, the method of the invention comprises a prior step of transformation of host cells with a plasmid as defined above, followed by culturing of the transformed cells, enabling suitable amounts of plasmid to be obtained. Excision by site-specific recombinations is then carried out by bringing into contact with the recombinase under the conditions defined above (FIG. 2). As stated above, in this embodiment, the site-specific recombination may be carried out in vivo. (that is to say in the host cell) or in vitro (that is to say on a plasmid preparation).

According to a preferred embodiment, the DNA molecules of the invention are hence obtained from a replicative vector, by excision of the non-therapeutic portion carrying, in particular, the origin of replication and the marker gene, by site-specific recombination.

According to another embodiment, the method of the invention comprises the excision of the DNA molecules from the genome of the host cell by site-specific recombination. This embodiment is based more especially on the construction of cell hosts comprising, inserted into their genome, one or more copies of a cassette comprising the gene of interest flanked by the sequences permitting recombination (FIG. 1). Different techniques may be used for insertion of the cassette of the invention into the genome of the host cell. In particular, insertion at several distinct points of the genome may be obtained by using integrative vectors. In this connection, different transposition systems such as, in particular, the miniMu system or defective transposons such as Tn10 derivatives, for example, may be used (Kleckner et al., Methods Enzymol. 204 (1991) 139; Groisman E., Methods Enzymol. 204 (1991) 180). The insertion may also be carried out by homologous recombination, enabling a cassette containing two recombination sequences in the direct orientation flanking one or more genes of interest to be integrated in the genome of the bacterium. This process may, in addition, be reproduced as many times as desired so as to have the largest possible number of copies per cell. Another technique also consists in using an in vivo amplification

system using recombination, as described in Labarre et al. (Labarre J., O. Reyes, Guyonvarch, and G. Leblon. 1993. Gene replacement, integration, and amplification at the gdhA locus of *Corynebacterium glutamicum*. J. Bacteriol. 175:1001–107), so as to augment from one copy of the cassette to a much larger number.

A preferred technique consists in the use of miniMu. To this end, miniMu derivatives are constructed comprising a resistance marker, the functions required in cis for their transposition and a cassette containing two recombination sequences in the direct orientation flanking the gene or genes of interest. These miniMus are advantageously placed at several points of the genome using a resistance marker (kanamycin, for example) enabling several copies per genome to be selected (Groisman E. cited above). As described above, the host cell in question can also express inducibly a site-specific recombinase leading to the excision of the fragment flanked by the recombination sequences in the direct orientation. After excision, the minicircles may be purified by standard techniques.

This embodiment of the method of the invention is especially advantageous, since it leads to the generation of a single type of plasmid molecule: the minicircle of the invention. The cells do not contain, in effect, any other episomal plasmid, as is the case during production from a plasmid (FIGS. 1 and 2).

Another subject of the invention also lies in a modified host cell comprising, inserted into its genome, one or more copies of a recombinant DNA as defined above.

The invention also relates to any recombinant cell containing a plasmid as defined above. These cells are obtained by any technique known to a person skilled in the art enabling a DNA to be introduced into a given cell. Such a technique can be, in particular, transformation, electroporation, conjugation, protoplast fusion or any other technique known to a person skilled in the art. As regards transformation, different protocols have been described in the prior art. In particular, cell transformation may be carried out by treating whole cells in the presence of lithium acetate and polyethylene glycol according to the technique described by Ito et al. (J. Bacteriol. 153 (1983) 163–168), or in the presence of ethylene glycol and dimethyl sulphoxide according to the technique of Durrens et al. (Curr. Genet. 18 (1990) 7). An alternative protocol has also been described in Patent Application EP 361,991. As regards electroporation, this may be carried out according to Becker and Guarentte (in: Methods in Enzymology Vol194 (1991) 182).

The method according to the invention may be carried out in any type of cell host. Such hosts can be, in particular, bacteria or eukaryotic cells (yeasts, animal cells, plant cells), and the like. Among bacteria, *E.coli, B. subtilis*, Streptomyces, Pseudomonas (*P. putida, P. aeruginosa*), *Rhizobium meliloti, Agrobacterium tumefaciens, Staphylococcus aureus, Streptomyces prisinaespiralis, Enterococcus faecium* or Clostridium, and the like, may be mentioned more preferentially. Among bacteria, it is preferable to use *E.coli*. Among yeasts, Kluyveromyces, Saccharomyces, Pichia, Hansenula, and the like, may be mentioned. Among mammalian animal cells, CHO, COS, NIH3T3, and the like, cells may be mentioned.

In accordance with the host used, the plasmid according to the invention is adapted by a person skilled in the art to permit its replication. In particular, the origin of replication and the marker gene are chosen in accordance with the host cell selected.

The marker gene may be a resistance gene, in particular for resistance to an antibiotic (ampicillin, kanamycin, geneticin, hygromycin, and the like), or any gene endowing the cell with a function which it no longer possesses (for example a gene which has been deleted on the chromosome or rendered inactive), the gene on the plasmid reestablishing this function.

In a particular embodiment, the method of the invention comprises an additional step of purification of the minicircle.

In this connection, the minicircle may be purified by standard techniques of plasmid DNA purification, since it is supercoiled like plasmid DNA. These techniques comprise, inter alia, purification on a cesium chloride density gradient in the presence of ethidium bromide, or alternatively the use of anion exchange columns (Maniatis et al., 1989). In addition, if the plasmid DNA corresponding to the non-therapeutic portions (origin of replication and selectable marker in particular) is considered to be present in an excessively large amount, it is also possible, after or before the purification, to use one or more restriction enzymes which will digest the plasmid and not the minicircle, enabling them to be separated by techniques that separate supercoiled DNA from linear DNA, such as a cesium chloride density gradient in the presence of ethidium bromide (Maniatis et al., 1989).

In addition, the present invention also describes an improved method for the purification of minicircles. This method enables minicircles of very great purity to be obtained in large yields in a single step. This improved method is based on the interaction between a double-stranded sequence present in the minicircle and a specific ligand. The ligand can be of various natures, and in particular protein, chemical or nucleic acid in nature. It is preferably a ligand of the nucleic acid type, and in particular an oligonucleotide, optionally chemically modified, capable of forming by hybridization a triple helix with the specific sequence present in the DNA molecule of the invention. It was, in effect, shown that some oligonucleotides were capable of specifically forming triple helices with double-stranded DNA sequences (Hélène et al., Biochim. Biophys. Acta 1049 (1990) 99; see also FR 94/15162 incorporated in the present application by reference).

In an especially advantageous variant, the DNA molecules of the invention hence contain, in addition, a sequence capable of interacting specifically with a ligand (FIG. 3). Preferably, it is a sequence capable of forming, by hybridization, a triple helix with a specific oligonucleotide. This sequence may be positioned at any site of the DNA molecule of the invention, provided it does not affect the functionality of the gene of interest. This sequence is also present in the genetic constructions of the invention (plasmids, cassettes), in the portion containing the gene of interest (see, in particular, the plasmid pXL2650). Preferably, the specific sequence present in the DNA molecule of the invention comprises between 5 and 30 base pairs.

The oligonucleotides used for carrying out the method according to the invention can contain the following bases:

- thymidine (T), which is capable of forming triplets with A.T doublets of double-stranded DNA (Rajagopal et al., Biochem 28 (1989) 7859);
- adenine (A), which is capable of forming triplets with A.T doublets of double-stranded DNA;
- guanine (G), which is capable of forming triplets with G.C doublets of doubled-stranded DNA;
- protonated cytosine (C+), which is capable of forming triplets with G.C doublets of doubled-stranded DNA (Rajagopal et al., cited above).

Preferably, the oligonucleotide used comprises a homopyrimidine sequence containing cytosines, and the specific sequence present in the DNA molecule is a homopurine-homopyrimidine sequence. The presence of cytosines makes it possible to have a triple helix which is stable at acid pH where the cytosines are protonated, and destabilized at alkaline pH where the cytosines are neutralized.

To permit the formation of a triple helix by hybridization, it is important for the oligonucleotide and the specific sequence present in the DNA molecule of the invention to be complementary. In this connection, to obtain the best yields and best selectivity, an oligonucleotide and a specific sequence which are fully complementary are used in the method of the invention. Possible combinations are, in particular, a poly(CTT) oligonucleotide and a poly(GAA) specific sequence. By way of example, there may be mentioned the oligonucleotide of sequence GAGGCTTCTTCT-TCTTCTTCTTCTT (SEQ ID No. 5), in which the bases GAGG do not form a triple helix but enable the oligonucleotide to be spaced apart from the coupling arm.

It is understood, however, that some mismatches may be tolerated, provided they do not lead to too great a loss of affinity. The oligonucleotide used may be natural (composed of unmodified natural bases) or chemically modified. In particular, the oligonucleotide may advantageously possess some chemical modifications enabling its resistance or its protection against nucleases, or its affinity for the specific sequence, to be increased.

Thus, the oligonucleotide may be rendered more resistant to nucleases by modification of the skeleton (e.g. methylphosphonates, phosphorothiates, phosphotriester, phosphoramidate, and the like). Another type of modification has as its objective, more especially, to improve the interaction and/or the affinity between the oligonucleotide and the specific sequence. In particular, a thoroughly advantageous modification according to the invention consists in methylating the cytosines of the oligonucleotide. The oligonucleotide thus methylated displays the noteworthy property of forming a stable triple helix with the specific sequence at neutral pH. Hence it makes it possible to work at higher pH values than the oligonucleotides of the prior art, that is to say at pH values where the risks of degradation of the plasmid DNA are lower.

The length of the oligonucleotide used in the method of the invention is at least 3 bases, and preferably between 5 and 30. An oligonucleotide of length greater than 10 bases is advantageously used. The length may be adapted to each individual case by a person skilled in the art in accordance with the desired selectivity and stability of the interaction.

The oligonucleotides according to the invention may be synthesized by any known technique. In particular, they may be prepared by means of nucleic acid synthesizers. It is quite obvious that any other method known to a person skilled in the art may be used.

To carry out the method of the invention, the specific ligand (protein, nucleic acid, and the like) may be grafted or otherwise onto a support. Different types of supports may be used for this purpose, such as, in particular, functionalized chromatography supports, in bulk form or prepacked in columns, functionalized plastic surfaces or functionalized latex beads, magnetic or otherwise. Chromatography supports are preferably used. By way of example, the chromatography supports which may be used are agatose, acrylamide or dextran, as well as their derivatives (such as Sephadex, Sepharose, Superose, etc.), polymers such as poly(styrenedivinylbenzene), or grafted or ungrafted silica, for example. The chromatography columns can function in the diffusion or perfusion mode.

To permit its covalent coupling to the support, the ligand is generally functionalized. In the case of an oligonucleotide, this may be modified, for example, with a terminal thiol, amine or carboxyl group at the 5' or 3' position. In particular, the addition of a thiol, amine or carboxyl group makes it possible, for example, to couple the oligonucleotide to a support carrying disulphide, maleimide, amine, carboxyl, ester, epoxide, cyanogen bromide or aldehyde functions. These couplings form by the establishment of disulphide, thioether, ester, amide or amine links between the oligonucleotide and the support. Any other method known to a person skilled in the art may be used, such as bifunctional coupling reagents, for example.

Moreover, to improve the activity of the coupled oligomucleotide, it may be advantageous to perform the coupling by means of an "arm". Use of an arm makes it possible, in effect, to bind the oligonucleotide at a chosen distance from the support, enabling its conditions of interaction with the DNA molecule of the invention to be improved. The arm advantageously consists of nucleotide bases that do not interfere with the hybridization. Thus, the arm may comprise purine bases. By way of example, the arm may comprise the sequence GAGG.

The DNA molecules according to the invention may be used in any application of vaccination or of gene and cell therapy, for the transfer of a gene to a body, a tissue or a given cell. In particular, they may be used for a direct administration in vivo, or for the modification of cells in vitro or ex vivo with a view to their implantation in a patient. In this connection, the molecules according to the invention may be used as they are (in the form of naked DNA), or in combination with different synthetic or natural, chemical and/or biochemical vectors. The latter can be, in particular, cations (calcium phosphate, DEAE-dextran, etc.) which act by forming precipitates with DNA, which precipitates can be "phagocytosed" by the cells. They can also be liposomes in which the DNA molecule is incorporated and which fuse with the plasma membrane. Synthetic gene transfer vectors are generally lipids or cationic polymers which complex DNA and form a particle therewith carrying positive surface charges. These particles are capable of interacting with the negative charges of the cell membrane and then of crossing the latter. DOGS (Transfectam™) or DOTMA (Lipofectin™) may be mentioned as examples of such vectors. Chimeric proteins have also been developed: they consist of a polycationic portion which condenses DNA, linked to a ligand which binds to a membrane receptor and carries the complex into the cells by endocytosis. The DNA molecules according to the invention may also be used for gene transfer into cells by physical transfection techniques such as bombardment, electroporation, and the like. In addition, prior to their therapeutic use, the molecules of the invention may optionally be linearized, for example by enzymatic cleavage.

In this connection, another subject of the present invention relates to any pharmaceutical composition comprising at least one DNA molecule as defined above. This molecule may be naked or combined with a chemical and/or biochemical transfection vector. The pharmaceutical compositions according to the invention may be formulated with a view to topical, oral, parenteral, intranasal, intravenous, intramuscular, subcutaneous, intra-ocular, transdermal, and the like, administration. Preferably, the DNA molecule is used in an injectable form or by application. It may be mixed with any pharmaceutically acceptable vehicle for an injectable formulation, in particular for a direct injection at the site to be treated. The compositions can be, in particular, in the

form of isotonic sterile solutions, or of dry, in particular lyophilized compositions which, on addition of sterilized water or physiological saline as appropriate, enable injectable solutions to be made up. Diluted Tris or PBS buffers in glucose or sodium chloride may be used in particular. A direct injection of the nucleic acid into the affected region of the patient is advantageous, since it enables the therapeutic effect to be concentrated in the tissues affected. The doses of nucleic acid used may be adapted in accordance with different parameters, and in particular in accordance with the gene, the vector, the mode of administration used, the pathology in question or alternatively the desired treatment period.

The DNA molecules of the invention may contain one or more genes of interest, that is to say one or more nucleic acids (cDNA, gDNA, synthetic or semi-synthetic DNA, and the like) whose transcription and, where appropriate, translation in the target cell generate products of therapeutic, vaccinal, agricultural or veterinary value.

Among the genes of therapeutic value, there may be mentioned, more especially, the genes coding for enzymes, blood derivatives, hormones, lymphokines, namely interleukins, interferons, TNF, and the like (FR 92/03120), growth factors, neurotransmitters or their precursors or synthetic enzymes, trophic factors, namely BDNF, CNTF, NGF, IGF, GMF, aFGF, bFGF, NT3, NT5, and the like; apoliproproteins, namely ApoAI, ApoAIV, ApoE, and the like (FR 93/05125), dystrophin or a minidystrophin (FR 91/11947), tumour suppressive genes, namely p53, Rb, Rap1A, DCC, k-rev, and the like (FR 93/04745), genes coding for factors involved in coagulation, namely factors VII, VIII, IX, and the like, suicide genes, namely thymidine kinase, cytosine deaminase, and the like; or alternatively all or part of a natural or artificial immunoglobulin (Fab, ScFv, and the like), a ligand RNA (WO91/19813), and the like. The therapeutic gene can also be an antisense gene or sequence whose expression in the target cell enables gene expression or the transcription of cellular mRNAs to be controlled. Such sequences can, for example, be transcribed in the target cell into RNAs complementary to cellular mRNAs, and can thus block their translation into protein, according to the technique described in Patent EP 140,308.

The gene of interest can also be a vaccinating gene, that is to say a gene coding for an antigenic peptide, capable of generating an immune response in man or animals for the purpose of vaccine production. Such antigenic peptides can be, in particular, those specific to the Epstein-Barr virus, the HIV virus, the hepatitis B virus (EP 185,573) or the pseudorabies virus, or alternatively tumour-specific peptides (EP 259,212).

Generally, in the plasmids and molecules of the invention, the gene of therapeutic, vaccinal, agricultural or veterinary value also contains a transcription promoter region which is functional in the target cell or body (i.e. mammals), as well as a region located at the 3' end and which specifies a transcription termination signal and a polyadenylation site (expression cassette). As regards the promoter region, this can be a promoter region naturally responsible for the expression of the gene in question when the latter is capable of functioning in the cell or body in question. The promoter regions can also be those of different origin (responsible for the expression of other proteins, or even synthetic promoters). In particular, the promoter sequences can be from eukaryotic or viral genes. For example, they can be promoter sequences originating from the genome of the target cell. Among eukaryotic promoters, it is possible to use any promoter or derived sequence that stimulates or represses the transcription of a gene, specifically or otherwise, inducibly or otherwise, strongly or weakly. They can be, in particular, ubiquitous promoters (promoter of the HPRT, PGK, α-actin, tubulin, and the like, genes), promoters of intermediate filaments (promoter of the GFAP, desmin, vimentin, neurofilament, keratin, and the like, genes), promoters of therapeutic genes (for example the promoter of the MDR, CFTR, factor VIII, ApoAI, and the like, genes), tissue-specific promoters (promoter of the pyruvate kinase gene, villin gene, gene for intestinal fatty acid binding protein, gene for α-actin of smooth muscle, and the like) or alternatively promoters that respond to a stimulus (steroid hormone receptor, retinoic acid receptor, and the like). Similarly, the promoter sequences may be those originating from the genome of a virus, such as, for example, the promoters of the adenovirus E1A and MLP genes, the CMV early promoter or alternatively the RSV LTR promoter, and the like. In addition, these promoter regions may be modified by the addition of activator or regulator sequences or sequences permitting a tissue-specific or -preponderant expression.

Moreover, the gene of interest can also contain a signal sequence directing the synthesized product into the pathways of secretion of the target cell. This signal sequence can be the natural signal sequence of the product synthesized, but it can also be any other functional signal sequence, or an artificial signal sequence.

Depending on the gene of interest, the DNA molecules of the invention may be used for the treatment or prevention of a large number of pathologies, including genetic disorders (dystrophy, cystic fibrosis, and the like), neurodegenerative diseases (Alzheimer's, Parkinson's, ALS, and the like), cancers, pathologies associated with disorders of coagulation or with dysliproproteinaemias, pathologies associated with viral infections (hepatitis, AIDS, and the like), or in the agricultural and veterinary fields, and the like.

The present invention will be described more completely by means of the examples which follow, which are to be regarded as illustrative and non-limiting.

BRIEF DESCRIPTION OF THE FIGURES

FIG. 1: Production of a minicircle from a cassette integrated in the genome.

FIG. 2: Production of a minicircle from a plasmid.

FIG. 3: Production of a minicircle containing a sequence specific to a ligand.

FIG. 4: Construction of pXL2649. Ori: Origin of replication; Kan^r: Marker gene conferring resistance to kanamycin; Amp^r: Marker gene conferring resistance to ampicillin; galK: Galactosidase gene of *E.coli*; Plac: Promoter of the lactose operon.

FIG. 5: Luciferase activity obtained after transfection of NIH3T3 mouse fibroblasts with plasmid pXL2650, the minicircle generated from plasmid pXL2650 and PGL2-Control (Promega, Biotech). The transfection was carried out under the following conditions: 0.5 mg of DNA per well, 50,000 cells per well. The lipofectant used is RPR 115335. The result is recorded in RLU per microgram of proteins as a function of the lipofectant/DNA charge ratio.

FIG. 6: Construction of the plasmid pXL2793. This plasmid generates, after recombination, a minicircle containing a synthetic homopurine-homopyrimidine sequence and the luciferase cassette of pXL2727.

FIG. 7: Well 1 corresponds to the SalI digestion of the fraction eluted after purification with a triple-helix column. Well 2 corresponds to the XmnI digestion of the fraction

eluted after purification with a triple-helix column. Well **3** corresponds to the undigested fraction eluted after purification with a triple-helix column. Well **4** corresponds to uninduced, undigested plasmid pXL2793. Wells **5** and **6** correspond, respectively, to the linear DNA and supercoiled DNA size markers.

FIG. **8**: Diagrammatic description of the construction of the plasmid pXL2776.

FIG. **9**: Diagrammatic description of the constructions of the plasmids pXL2777 and pXL2960.

FIG. **10**: Action of the integrase of bacteriophage l in *E. coli* on plasmids pXL2777 and pXL2960. M: linear DNA or supercoiled DNA 1 kb molecular weight marker. N.I.: not induced. I: induced. N.D.: not digested.

FIG. **11**: Kinetics of recombination of the integrase of bacteriophage 1 in *E. coli* on plasmids pXL2777 and pXL2960. 2': 2 minutes. O/N: 14 hours. M: linear DNA or supercoiled DNA 1 kg molecular weight marker. N.I.: not induced. I: induced. N.D.: not digested.

General techniques of cloning and molecular biology.

The standard methods of molecular biology, such as centrifugation of plasmid DNA in a cesium chloride-ethidium bromide gradient, digestion with restriction enzymes, gel electrophoresis, electroelution of DNA fragments from agarose gels, transformation in *E.coli*, precipitation of nucleic acids, and the like, are described in the literature (Maniatis et al., 1989, Ausubel et al., 1987). Nucleotide sequences were determined by the chain termination method according to the protocol already put forward (Ausubel et al., 1987).

Restriction enzymes were supplied by New-England Biolabs (Biolabs), Bethesda Research Laboratories (BRL) or Amersham Ltd. (Amersham).

To carry out ligation, DNA fragments are separated according to their size on 0.7 % agarose or 8% acrylamide gels, purified by electrophoresis and then electroelution, extracted with phenol, precipitated with ethanol and then incubated in a buffer comprising 50 mM Tris-HCl, pH 7.4, 10 mM $MgCl_2$, 10 mM, DTT, 2 mM ATP in the presence of phage T4 DNA ligase (Biolabs). Oligo-nucleotides are synthesized using phosphoramidite chemistry with the latter derivatives protected at the b position by a cyanoethyl group (Sinha et al., 1984, Giles 1985), with the Biosearch 8600 automatic DNA synthesizer, using the manufacturer's recommendations.

The ligated DNAs are used to transform the following strains rendered competent†: *E.coli* MC1060 [(LacIOPZYA)X74, galU, galK, strA^r, hsdR] (Casadaban et al., 1983); HB101 [hsdS20, supE44, recA13, ara-14, proA2, lacY1, galK2, rpsL20, xyl-5, mtl-1, λ−, F−] (Maniatis et al., 1989); and DH5α [endA1 hsdR17 supE44 thi-1 recA1 gyrA96 relA1 λ-Φ80 dlacZΔM15] for the plasmids.

LB and 2XTY culture media are used for the bacteriological part (Maniatis et al., 1989).

Plasmid DNAs are purified according to the alkaline lysis technique (Maniatis et al., 1989).

Definition of the terms employed and abbreviations.

Recombinant DNA: set of techniques which make it possible either to combine, within the same microorganism, DNA sequences which are not naturally combined, or to mutagenize a DNA fragment specifically.

ATP: adenosine 5'-triphosphate

BSA: bovine serum albumin

PBS: 10 mM phosphate buffer, 150 mM NaCl, pH 7.4

dNTP: 2'-deoxyribonucleoside 5'-triphosphates

DTT: dithiothreitol

kb: kilobases

bp: base pairs

EXAMPLE 1

Construction of a Plasmid Carrying the attP and attB Sequences of the Bacteriophage, in Repeated Direct Orientations.

The plasmid pNH16a was used as starting material, inasmuch as it already contains a fragment of bacteriophage λ carrying the attP sequence (Hasan and Szybalski, 1987). This plasmid was digested with EcoRI. Oligonucleotides which contain the attB sequence (Landy, 1989) were synthesized. They have the following sequence:

Oligonucleotide 5476 (SEQ ID No.1)
5'-AATTGTGAAGCCTGCTTTTTTATACTAAC TTGAGCGG-3'

Oligonucleotide 5477 (SEQ ID No.2)
5'-AATTCCGCTCAAGTTAGTATAAAAAAGCA GGCTTCAC-3'

They were hybridized to re-form the attB sequence and then ligated at the EcoRI site of the 4.2-kb EcoRI fragment of pNH16a (Hasan and Szybalski, 1987). After transformation of DH5α, a recombinant clone was retained. The plasmid thereby constructed was designated pXL2648 (see FIG. **4**). This plasmid contains the attP and attB sequences of the bacteriophage in the direct orientation. Under the action of the integrase of the bacteriophage (Int protein), there should be excision of the sequences lying between the two att sites. This results in separation of the material inserted between the two att sequences from the origin of replication and from the resistance marker of the plasmid, which are positioned on the outside.

EXAMPLE 2

Obtaining a Minicircle in vivo in *E.coli*.

A cassette for resistance to kanamycin was cloned at the EcoRI site of plasmid pXL2648 (FIG. **4**). This cassette originates from the plasmid pUC4KIXX (Pharmacia Biotech.). For this purpose, 10 μg of plasmid pUC4KIXX were digested with EcoRI and then separated by agarose gel electrophoresis; the 1.6-kb fragment containing the kanamycin resistance marker was purified by electro-elution; it was then ligated to plasmid pXL2648 linearized with EcoRI. The recombinant clones were selected after transformation into *E.coli* DH5α and selection for resistance to kanamycin. The expected restriction profile was observed on one clone; this plasmid clone was designated pXL2649 (FIG. **4**). This plasmid was introduced by transformation into two *E.coli* strains:

D1210 [hsdS20, supE44, recA13, ara-14, proA2, lacY1, galK2, rpsL20, xyl-5, mtl-1, λ$^-$, F−, lacIq] (Sadler et al., 1980).

D1210HP, which corresponds to DH1210 lysogenized with the phage xis$^-$ (Xis$^-$ Kil$^-$) cI857 (Podjaska et al., 1985). The D1210HP strain [supE44 ara-14 galK2 Δ(gpt-proA)62 rpsL20 xy15 mtl1 recA13 Δ(mcrC-mrr) hsdS lacI^q] (λ[cI857 xis$^-$ ki1$^-$]), accession number I-2314, was deposited on Sep. 15, 1999 with the Collection National de Cultures de Microorganisms (CNCM), Institut Pasteur, 25 rue du Docteur Roux, F-75724 Paris Cédex 15, FRANCE.

The transformants were selected at 30° C. on 2XTY medium with kanamycin (50 mg/l). After reisolation on selective medium, the strains were inoculated into 5 ml of L

medium supplemented with kanamycin (50 mg/l). After 16 h of incubation at 30° C. with agitation (5 cm of rotational amplitude), the cultures were diluted to 1/100 in 100 ml of the same medium. These cultures were incubated under the same conditions until an OD_{610} of 0.3 was reached. At this point, half of the culture was removed and then incubated for 10 min at 42° C. to induce the lytic cycle of the phage, hence the expression of the integrase. After this incubation, the cultures were transferred again to 30° C. and then incubated for 1 h under these conditions. Next, culturing was stopped and minipreparations of plasmid DNA were produced. Irrespective of the conditions, in the strain D1210, the agarose gel electrophoresis profile of the undigested plasmid DNA of plasmid pXL2649 is unchanged, as is also the case in the strain D1210HP which has not been thermally induced. On the contrary, in D1210HP which has been incubated for 10 min at 42° C. and then cultured for 1 hour at 30° C., it is found that there is no longer a plasmid, but two circular DNA molecules: one of low molecular weight, migrating faster and containing an EcoRI site; and one of higher molecular weight, containing a unique BglI site, as expected. Hence there has indeed been excision of the sequences present between the two att sequences, and generation of a minicircle bereft of any origin of replication. This supercoiled circular DNA not carrying an origin of replication is termed a minicircle. This name takes, in effect, better account of the circular nature of the molecule. The starting plasmid pXL2649 is present, but it represents approximately 10% of the plasmid which has excised the sequences flanked by att.

The minicircle may then be purified by standard techniques of plasmid DNA purification, since it is supercoiled like plasmid DNA. These techniques comprise, inter alia, purification on a cesium chloride density gradient in the presence of ethidium bromide, or alternatively the use of anion exchange columns (Maniatis et al., 1989). In addition, if the plasmid DNA corresponding to the origin of replication and to the selectable marker is considered to be present in an excessively large amount, it is always possible, after purification, to use one or more restriction enzymes which will digest the plasmid and not the minicircle, enabling them to be separated by techniques that separate supercoiled DNA from linear DNA, such as in a cesium chloride density gradient in the presence of ethidium bromide (Maniatis et al., 1989).

EXAMPLE 3

Obtaining a Minicircle Containing a Cassette for the Expression of Luciferase.

In order to test the use of these minicircles in vivo, a reporter gene with the sequences required for its expression was cloned into plasmid pXL2649 (see Example 2). This was done using, more especially, a 3150-bp BglII-BamHI cassette originating from pGL2-Control (Promega Biotech). This cassette contains the SV40 early promoter, the enhancer of the SV40 early promoter, the luciferase gene of *Photinus pyralis* and a polyadenylation site derived from SV40. The 3150-bp BglII-BamHI fragment was cloned at the BamHI site of pXL2649 digested with BamHI so as to replace the cassette for resistance to kanamycin by the cassette for the expression of luciferase from pGL2-control. The plasmid thus constructed was called pXL2650. In this plasmid, the attP and attB sites flank the cassette for the expression of luciferase. Site-specific recombination enables only the sequences required for the expression of luciferase together with the luciferase gene to be excised. This recombination may be carried out exactly as described in Example 2. A minicircle such as plasmid pXL2650 may be used thereafter in in vivo or in vitro transfection experiments.

A 1-liter culture of the strain D1210HP pXL2650 in 2XTY medium supplemented with ampicillin (50 mg/ml) was set up at 30° C. At an OD_{610} equal to 0.3, the culture was transferred to 42° C. for 20 min, then replaced for 20 min at 30° C. The episomal DNA was prepared by the clear lysate technique (Maniatis et al., 1989), followed by a cesium chloride density gradient supplemented with ethidium bromide (Maniatis et al., 1989), then by an extraction of the ethidium bromide with isopropanol and by a dialysis. This DNA was shown to contain the minicircle. 100 μg of this preparation were digested with PstI, and the hydrolysate was then subjected to a cesium chloride density gradient supplemented with ethidium bromide (Maniatis et al., 1989). An identical result is obtained when the preparation is digested jointly with AlwNI and XmnI. The supercoiled form was recovered and, after removal of the ethidium bromide (Maniatis et al.), it was found to correspond only to the minicircle, lacking an origin of replication and any marker gene. This minicircle preparation may be used for in vitro and in vivo transfection experiments.

EXAMPLE 4

In vitro Transfection of Mammalian Cells, and More Especially of Human Cells, with a Minicircle.

The minicircle DNA containing the luciferase gene of *Photinus pyralis* as described in Example 3, that is to say corresponding to the minicircle generated from plasmid pXL2650, is diluted in 150 mM NaCl and mixed with a transfectant. It is possible to use various commercial transfectants, such as dioctadecylamidoglycylspermine (DOGS, Transfectam™, Promega), Lipofectin™ (Gibco-BRL), and the like, in different positive/negative charge ratios. By way of illustration, the transfecting agent was used in charge ratios greater than or equal to 3. The mixture is vortexed, left for 10 minutes at room temperature, diluted in culture medium without fetal calf serum, and then added to the cells in the proportion of 2 μg of DNA per culture well. The cells used are Caco-2, derived from a human colon adenocarcinoma, cultured according to a protocol described (Wils et al., 1994) and inoculated on the day before the experiment into 48-well culture plates in the proportion of 50,000 cells/well. After two hours at 37° C., 10% v/v of fetal calf serum is added and the cells are incubated for 24 hours at 37° C. in the presence of 5% CO_2. The cells are washed twice with PBS and the luciferase activity is measured according to the protocol described (such as the Promega kit). It is possible to use other lines (fibroblasts, lymphocytes, etc.) originating from different species, or alternatively cells taken from an individual (fibroblasts, keratinocytes, lymphocytes, etc.) and which will be reinjected into him or her after transfection.

EXAMPLE 5

In vitro Transfection of NIH 3T3 Cells.

The minicircle DNA containing the luciferase gene of *Photinus pyralis*, as described in Example 3, that is to say corresponding to the minicircle generated from plasmid pXL2650, was transfected in vitro into mammalian cells; pXL2650 and PGL2-Control (Promega Biotech.), which contain the same expression cassette, were used as control. The cells used are NIH 3T3 mouse fibroblasts, inoculated on the day before the experiment into 24-well culture plates in the proportion of 50,000 cells per well. The plasmid is diluted in 150 mM NaCl and mixed with the lipofectant RPR115335. However, it is possible to use various other commercial agents such as dioctadecylaminoglycylspermine (DOGS, Transfectam™, Promega) (Demeneix et al., Int. J. Dev. Biol. 35 (1991) 481), Lipofectin™ (Gibco-BRL)

(Fegner et al., Proc. Natl. Acad. Sci. USA 84 (1987) 7413), and the like. A positive charge of the lipofectant/negative charge of the DNA ratio equal to or greater than 3 is used. The mixture is vortexed, left for ten minutes at room temperature, diluted in medium without fetal calf serum, and then added to the cells in the proportion of 0.5 mg of DNA per culture well. After two hours at 37° C., 10% by volume of fetal calf serum is added and the cells are incubated for 48 hours at 37° C. in the presence of 5% CO_2. The cells are washed twice with PBS and the luciferase activity is measured according to the protocol described (Promega kit, Promega Corp. Madison, Wis.), on a Lumat LB9501 luminometer (EG and G Berthold, Evry). The transfection results corresponding to the conditions which have just been stated are presented in FIG. 5. They show unambiguously that the minicircle has the same transfection properties as plasmids possessing an origin of replication. Thus these minicircles could be used in the same way as standard plasmids in gene therapy applications.

EXAMPLE 6

Affinity Purification of a Minicircle Using a Triple-helix Interaction.

This example describes a method of purification of a minicircle according to the invention from a mixture containing the plasmid form which has excised it, by triple-helix type interactions which will take place with a synthetic DNA sequence carried by the minicircle to be purified. This example demonstrates how the technology of purification by triple-helix formation may be used to separate a minicircle from a plasmid form which has excised it.

6-1. Obtaining a Minicircle Containing a Synthetic Homopurine-homopyrimidine Sequence

6-1.1. Insertion of a homopurine-homopyrimidine sequence into plasmid pXL2650

Plasmid pXL2650 possesses a unique BamHI site immediately after the cassette containing the luciferase gene of *Photinus pyralis*. This unique site was used to clone the following two oligonucleotides:

4957 (SEQ ID No.3) 5'-GATCCGAAGAAGAAGAAGAAGAAGAAG AAGAAGAAGAAGAAGAAGAAGAAGAAGAAC-3'

4958 (SEQ ID No.4) 5'-GATCGTTCTTCTTCTTCTTCTTCTTCTTCT TCTTCTTCTTCTTCTTCTTCTTCTTCG-3'

These oligonucleotides, when hybridized and cloned into plasmid pXL2650, introduce a homopurine-homopyrimidine sequence $(GAA)_{17}$, as described above.

To carry out this cloning, the oligonucleotides were first hybridized in the following manner. One μg of each of these two oligonucleotides were placed together in 40 ml of a final buffer comprising 50 mM Tris-HCl, pH 7.4, 10 mM $MgCl_2$. This mixture was heated to 95° C. and was then placed at room temperature so that the temperature would fall slowly. Ten ng of the mixture of hybridized oligonucleotides were ligated with 200 ng of plasmid pXL2650 linearized with BamHI, 30 ml of final. After ligation, an aliquot was transformed into DH5. The transformation mixtures were plated out on L medium supplemented with ampicillin (50 mg/l). Twenty-four clones were digested with PflMI and BamHI. One clone was found which had the size of the 950-bp PflMI-BamHI fragment increased by 50 bp. This clone was selected and designated pXL2651.

Plasmid pXL2651 was purified according to the Wizard Megaprep kit (Promega Corp., Madison, Wis.) according to the supplier's recommendations.

6-1.2. Insertion of a homopurine-homopyrimidine sequence into plasmid pXL2649

a) Insertion of new restriction sites on each side of the kanamycin cassette of pXL2649.

Plasmid pXL2649, as described in Example 2, was digested with EcoRI so as to take out the kanamycin cassette originating from plasmid pUC4KIXX (PharmaciaBiotech, Uppsala, Sweden). For this purpose, 5 mg of plasmid pXL2649 were digested with EcoRI. The 4.2-kb fragment was separated by agarose gel electrophoresis and purified by electroelution.

In addition, the plasmid pXL1571 was used. The latter was constructed from the plasmid pFR10 (Gene 25 (1983), 71–88), into which the 1.6-kb fragment originating from pUC4KIXX, corresponding to the kanamycin gene, was inserted at the SstI site. This cloning enabled 12 new restriction sites to be inserted on each side of the kanamycin gene.

Five micrograms of pXL1571 were dialysed with EcoRI. The 1.6-kb fragment corresponding to the kanamycin gene was separated by agarose gel electrophoresis and purified by electroelution. It was then ligated with the 4.2-kb EcoRI fragment of pXL2649. The recombinant clones were selected after transformation into *E. coli* DH5a and selection for resistance to kanamycin and to ampicillin. The expected restriction profile was observed on one clone; this plasmid clone was designated pXL2791.

b) Extraction of the kanamycin cassette from plasmid pXL2791

Plasmid pXL2791 was digested with SstI so as to take out the kanamycin cassette. The 4.2-kb fragment was separated by agarose gel electrophoresis and purified with the Jetsorb extraction gel kit (Genomed). It was then ligated. The recombinant clones were selected for resistance to ampicillin after transformation into *E. coli* DH5a. The expected restriction profile was observed on one clone. This plasmid clone was designated pXL2792. This clone comprises, inter alia, SalI and XmaI restriction sites between the attP and attB sites. c) Cloning of a homopurine-homopyrimidine sequence as well as of a cassette permitting the expression of luciferase between the two attP and attB sites of plasmid pXL2792

Plasmid pXL2727 was used. This plasmid, digested with XmaI and SalI, enables a fragment comprising the following to be taken out: the pCMV promoter, the luciferase gene of *Photinus pyralis*, a polyadenylation site derived from SV40 and a homopurine-homopyrimidine sequence. The latter was obtained after hybridization and cloning of the following two oligonucleotides:

6006: (SEQ ID No.16) 5'-GATCTGAAGAAGAAGAAGAAGAAGAAGA AGAAGAAGAAGAAGAAGAAGAAGAAGAAC TGCAGATCT-3'

6008: (SEQ ID No.17) 5'-GATCAGATCTGCAGTTCTTCTTCTTCTTCTT CTTCTTCTTCTTCTTCT TCTTCTTCTTCTTCTTCA-3'

The homopurine-homopyrimidine sequence present in pXL2727 was sequenced by the Sequenase Version 2.0 method (United States Biochemical Corporation). The result obtained shows that the homopurine-homopyrimidine sequence actually present in plasmid pXL2727 contains 10 repeats (GAA-CTT), and not 17 as the sequence of the oligonucleotides 6006 and 6008 suggested would be the case. The sequence actually present in plasmid pXL2727, read after sequencing on the strand corresponding to the oligonucleotide 6008, is as follows:

5'-GATCAGATCTGCAGTCTCTTCTTCTTCTT CTTCTTCTTCTTCT TCTTCTCTTCTCA-3' (SEQ ID No.18)

One microgram of pXL2727 was digested with XmaI and SalI. The 3.7-kb fragment was separated by agarose gel electrophoresis and purified with the Jetsorb extraction gel kit (Genomed). In addition, 1.7 mg of pXL2792 were digested with XmaI and SalI. The 4.2-kb fragment was separated on agarose gel, purified with the Jetsorb extraction gel kit (Genomed) and ligated with the 3.7-kb XmaI-SalI fragment of pXL2727. The recombinant clones were selected after transformation into *E. coli* DH5α and selection for resistance to ampicillin. The expected restriction profile was observed on one clone; this clone was designated pXL2793. Plasmid pXL2793 was purified using a caesium chloride density gradient according to a method already described (Maniatis et al., 1989).

6-2. Preparation of the Column Enabling Triple-helix Type Interactions with a Homopurine-homopyrimidine Sequence Present in the Minicircle to be Effected

The column was prepared in the following manner:

The column used is a 1-ml HiTrap column activated with NHS (N-hydroxysuccinimide, Pharmacia), connected to a peristaltic pump (flow rate<1 ml/min). The specific oligonucleotide used possesses an NH_2 group at the 5' end.

For plasmid pXL2651, its sequence is as follows:

5'-GAGGCTTCTTCTTCTTCTTCTTCTT-3' (SEQ ID No.5)

For plasmid pXL2793, its sequence is as follows (oligo 116418):

5'-CTTCTTCTTCTTCTTCTTCTT-3' (SEQ ID No. 19)

The buffers used are the following:

Coupling buffer: 0.2 M $NaHCO_3$, 0.5 M NaCl, pH 8.3.

Washing buffer:

Buffer A: 0.5 M ethanolamine, 0.5 M NaCl, pH 8.3.

Buffer B: 0.1 M acetate, 0.5 M NaCl, pH 4.

Fixing and eluting buffer:

Buffer F: 2 M NaCl, 0.2 M acetate, pH 4.5.

Buffer E: 1 M Tris-HCl, pH 9, 0.5 mM EDTA.

The column is prepared in the following manner:

The column is washed with 6 ml of 1 mM HCl, and the oligonucleotide diluted in the coupling buffer (50 nmol in 1 ml) is then applied to the column and left for 30 minutes at room temperature. The column is washed with 3 ml of coupling buffer, then with 6 ml of buffer A, followed by 6 ml of buffer B. The latter two buffers are applied three times in succession to the column. In this way, the oligonucleotide is linked covalently to the column via a CONH link. The column is stored at 4° C. in PBS, 0.1% NaN_3.

6-3. Purification of a Minicircle Containing a Synthetic Homopurine-homopyrimidine Sequence, by a Triple-helix Type Interaction

6-3.1. Purification of plasmid pXL2651

Plasmid pXL2651 was introduced into the strain D1210HP. This recombinant strain [D1210HP (pXL2651)] was cultured as described in Example 3 so as to generate the minicircle containing the luciferase gene of *Photinus pyralis*. Twenty ml of culture were removed and centrifuged. The cell pellet is taken up in 1.5 ml of 50 mM glucose, 25 mM Tris-HCl, pH 8, 10 mM EDTA. Lysis is carried out with 2 ml of 0.2 M NaOH, 1% SDS, and neutralization with 1.5 ml of 3 M potassium acetate, pH 5. The DNA is then precipitated with 3 ml of 2-propranol, and the pellet is taken up in 0.5 ml of 0.2 M sodium acetate, pH 5, 0.1 M NaCl and loaded onto an oligonucleotide column capable of forming triple-helix type interactions with poly(GAA) sequences contained in the minicircle, as described above. After the column has been washed beforehand with 6 ml of buffer F, the solution containing the minicircle to be purified is incubated, after being applied to the column, for two hours at room temperature. The column is washed with 10 ml of buffer F and elution is then carried out with buffer E.

Purified DNA corresponding to the minicircle is thereby obtained. The minicircle obtained, analysed by agarose gel electrophoresis and ethidium bromide staining, takes the form of a single band of supercoiled circular DNA. Less than 5% of starting plasmid pXL2651 is left in the preparation.

6-3.2. Purification of plasmid pXL2793

The 7.9-kb plasmid pXL2793 was introduced into the strain D1210HP. This recombinant strain was cultured as described in Example 3, so as to generate the 4-kb minicircle containing the luciferase gene of *Photinus pyralis* and a 3.9-kb plasmid. Two hundred ml of culture were removed and centrifuged. The cell pellet was treated with the Wizard Megaprep kit (Promega Corp., Madison, Wis.) according to the supplier's recommendations. The DNA was taken up in a final volume of 2 ml of 1 mM Tris, 1 mM EDTA, pH 8. Two hundred and fifty microliters of this plasmid sample were diluted with buffer F in a final volume of 2.5 ml. The column was washed beforehand with 6 ml of buffer F. The whole of the diluted sample was loaded onto an oligonucleotide column capable of forming triple-helix type interactions with poly(GAA) sequences contained in the minicircle, prepared as described above. After washing with 10 ml of buffer F, elution is carried out with buffer E. The eluted sample is recovered in 1-ml fractions.

By this method, purified DNA corresponding to the minicircle generated from pXL2793 is obtained. The DNA sample eluted from the column was analysed by agarose gel electrophoresis and ethidium bromide staining, and by enzyme restriction. For this purpose, the eluted fractions which were shown to contain DNA by assay at OD_{260} nm were dialysed for 24 hours against 1 mM Tris, 1 mM EDTA, then precipitated with isopropanol and taken up in 200 ml of H_2O. Fifteen microliters of the sample thereby obtained were digested with SalI, this restriction site being present in the minicircle and not in the 3.9-kb plasmid generated by the recombination, or with XmnI, this restriction site being present in the 3.9-kb plasmid generated by the recombination and not in the minicircle. The result obtained is presented in FIG. 7, showing that the minicircle has been purified of the recombinant plasmid.

EXAMPLE 7

In vivo Transfection of Mammalian Cells with a Minicircle

This example describes the transfer of a minicircle coding for the luciferase gene into the brain of newborn mice. The minicircle (30 μg) is diluted in sterile 150 mM NaCl to a concentration of 1 $\mu g/\mu l$. A synthetic transfectant such as dioctadecylamidoglycylspermine (DOGS) is then added in a positive/negative charge ratio less than or equal to 2. The mixture is vortexed, and 2 μg of DNA are injected into the cerebral cortex of anaesthetized newborn mice using a micromanipulator and a microsyringe. The brains are removed 48 hours later, homogenized and centrifuged and the supernatant is used for the assay of luciferase by the protocols described (such as the Promega kit).

EXAMPLE 8

Use of the par Locus of RK2 to Reduce the Presence of Minicircle or Miniplasmid Topoisomers

This example demonstrates the presence of topological forms derived i) from the plasmid possessing the attP and

attB sequences in the direct orientation, ii) from the minicircle or iii) from the miniplasmid, after the action of the integrase of bacteriophage l in *E. coli*. This example also shows that these topological or oligomeric forms may be resolved by using the par locus of RK2 (Gerlitz et al., 1990 J. Bacteriol. 172 p. 6194). In effect, this locus contains, in particular, the parA gene coding for a resolvase acting at the mrs (multimer resolution system) site (Eberl et al., 1994 Mol. Microbiol. 12 p. 131).

8-1. Construction of Plasmids pXL2777 and pXL2960

Plasmids pXL2777 and pXL2960 are derived from the vector pXL2776, and possess in common the minimal replicon of ColE1, the gene of the transposon Tn5 coding for resistance to kanamycin and the attP and attB sequences of bacteriophage l in the direct orientation. These plasmids differ in respect of the genes inserted between the attP and attB sequences, in particular pXL2777 contains the omegon cassette (coding for the gene for resistance to spectinomycin) whereas plasmid pXL2960 carries par locus of RK2.

8-1.1. Minimal vector pXL2658

The vector pXL2658 (2.513 kb) possesses the minimal replicon of ColE1 originating from pBluescript (ori) and the gene of the transposon Tn5 coding for resistance to kanamycin (Km) as selectable marker. After the BsaI end has been blunted by the action of the Klenow enzyme, the 1.15-kb BsaI-PvuII fragment of pBKS+ (obtained from Stratagene) was cloned with the 1.2-kb SmaI fragment of pUC4K1XX (obtained from Pharmacia) to generate the plasmid pXL2647. The oligo-nucleotides 5542 5'(AGC TTC TCG AGC TGC AGG ATA TCG AAT TCG GAT CCT CTA GAG CGG CCG CGA GCT CCG3' (SEQ ID No.20) and 5543 5'(AGC TGG AGC TCG CGG CCG CTC TAG AGG ATC CGA ATT CGA TAT CCT GCA GCT CGA GA)3' (SEQ ID No.21) were hybridized with one another and then cloned at the HindIII site of pXL2647; in this way pXL2658 is constructed. In this plasmid, the multiple cloning site is SstI, NotI, XbaI, BamHI, EcoRI, EcoRV, PstI, XhoI and HindIII between the origin of replication and the gene coding for resistance to kanamycin.

8-1.2. Vector pXL2776 containing the attP and attB sequences of phage l

The vector pXL2776 (2.93 kb) possesses the minimal replicon of ColE1 originating from pBluescript, the gene coding for resistance to kanamycin and the attP and attB sequences of bacteriophage l in the direct orientation, see FIG. 8. The 29-bp attB sequence (Mizuuchi et al., 1980 Proc. Natl. Acad. Sci. USA 77 p. 3220) was introduced between the SacI and HindIII restriction sites of pXL2658 after the sense oligonucleotide 6194 5'(ACT AGT GGC CAT GCA TCC GCT CAA GTT AGT ATA AAA AAG CAG GCT TCA G)3' (SEQ ID No.22) has been hybridized with the antisense oligonucleotide 6195 5'(AGC TCT GAA GCC TGC TTT TTT ATA CTA ACT TGA GCG GAT GCA TGG CCA CTA GTA GCT)3' (SEQ ID No.23) in such a way that the SacI and HindIII sites are no longer re-formed after cloning. This plasmid, the sequence of which was verified with respect to attB, is then digested with SpeI and NsiI in order to introduce in it the attP sequence flanked by the NsiI and XbaI restriction sites and thus to generate plasmid pXL2776. The attP sequence was obtained by PCR amplification using plasmid pXL2649 (described in Example 2) as template, the sense oligonucleotide 6190 5'(GCG TCT AGA ACA GTA TCG TGA TGA CAG AG)3' (SEQ ID No.24) and the antisense oligonucleotide 6191 5'(GCC AAG CTT AGC TTT GCA CTG GAT TGC GA)3' (SEQ ID No.25), and performing 30 cycles during which the hybridization temperature is 50° C. The PCR product digested at the XbaI and HindIII sites was cloned into the phage M13mpEH between the XbaI and HindIII sites. The amplified sequence is identical to the attP sequence described in Lambda II (edited by R. W. Hendrix, J. W. Roberts, F. W. Stahl, R. A. Weisberg; Cold Spring Harbor Laboratory 1983) between positions 27480 and 27863.

8-1.3. Plasmid pXL2777

Plasmid pXL2777 (6.9 kb) possesses the minimal replicon of ColE1 originating from pBluescript, the gene coding for resistance to kanamycin, the attP and attB sequences of bacteriophage l in the direct orientation and separated by the sacB gene coding for levansucrase of *B. subtilis* (P. Gay et al., 1983 J. Bacteriol. 153 p. 1424), and the Sp omegon coding for the gene for resistance to spectinomycin Sp and streptomycin Sm (P. Prentki et al., 1984 Gene 29 p. 303). The sacB-Sp cassette having EcoRV and NsiI cloning ends comes from the plasmid pXL2757 (FR95/01632) and was cloned between the EcoRV and NsiI sites of pXL2776 to form pXL2777.

8-1.4. Plasmid pXL2960

Plasmid pXL2960 (7.3 kb) possesses the minimal replicon of ColE1 originating from pBluescript, the gene coding for resistance to kanamycin and the attP and attB sequences of bacteriophage l in the direct orientation and separated by i) the sacB gene coding for levansucrase of *B. subtilis* (P. Gay et al., 1983 J. Bacteriol. 153 p. 1424) and ii) the par locus of RK2 (Gerlitz et al., 1990 J. Bacteriol. 172 p. 6194). The par cassette having BamHI ends comes from the plasmid pXL2433 (PCT/FR95/01178) and was introduced between the BamHI sites of pXL2777 to generate pXL2960.

8-2. Resolution of Minicircle or Miniplasmid Topoisomers

Plasmids pXL2777 and pXL2960 were introduced by transformation into *E. coli* strain D1210HP. The transformants were selected and assayed as described in Example 2, with the following modifications: the expression of the integrase was induced at 42° C. for 15 min when the optical density of the cells at 610 nm is 1.8, and the cells are then incubated at 30° C. for 30 min, see FIG. 9, or for a period varying from 2 minutes to 14 hours (O/N), see FIG. 10. The plasmid DNA originating from uninduced and induced cultures was then analysed on agarose gel before or after digestion with a restriction enzyme exclusive to the minicircle portion (EcoRI) or miniplasmid portion (BglII), see Figure Y, or after the action of DNA topoisomerase A or the gyrase of *E. coli*. The supercoiled dimer forms of minicircle or miniplasmid are clearly revealed by i) their molecular weight, ii) their linearization by the restriction enzyme, iii) their change in topology through the action of topoisomerase A (relaxed dimer) or of the gyrase (supersupercoiled dimer), iv) specific hybridization with an internal fragment peculiar to the minicircle or the miniplasmid. Other topological forms of higher molecular weights than that of the initial plasmid originate from the initial plasmid or the minicircle or the miniplasmid, since they disappear after digestion with the restriction enzyme exclusive to the minicircle portion (EcoRI) or miniplasmid portion (BglII). These forms are much less abundant with pXL2960 than with pXL2777 as initial plasmid, see FIG. 10. In particular, the dimer form of minicircle is present to a not insignificant extent with plasmid pXL2777, whereas it is invisible with plasmid pXL2960 when the cells are incubated for at least 30 min at 30° C., see FIGS. 9 and 10. It should be noted that minicircle dimers are observed at the beginning of the kinetic experiment with pXL2960 (2 to 10 min), and are thereafter resolved (after 30 min), see FIG. 10. Consequently, the par locus leads to a significant reduction

in the oligomeric/topological forms resulting from the action of the integrase of bacteriophage 1 in *E. coli* on plasmids containing the attP and attB sequences in the direct orientation.

IDENTIFICATION OF THE NUCLEOTIDE SEQUENCES

SEQ ID No.1: oligonucleotide 5476: 5'-AATTGTGAAGCCTGCTTTTTTATACTAACTTGAGCGG-3'

SEQ ID No.2: oligonucleotide 5477 5'-AATTCCGCTCAAGTTAGTATAAAAAAGCAGGCTTCAC-3'

SEQ ID No.3: oligonucleotide 4957: 5'-GATCCGAAGAGAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAC-3'

SEQ ID No.4: oligonucleotide 4958: 5'-GATCGTTCTTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCG-3'

SEQ ID No.5: oligonucleotide poly-CTT: 5'-GAGGCTTCTTCTTCTTCTTCTTCTT-3'

SEQ ID No.6: (attB sequence of phage lambda): 5'-CTGCTTTTTTATACTAACTTG-3'

SEQ ID No.7: (attP sequence of phage lambda): 5'-CAGCTTTTTTATACTAAGTTG-3'

SEQ ID No.8: (attB sequence of phage P22): 5'-CAGCGCATTCGTAATGCGAAG-3'

SEQ ID No.9: (attP sequence of phage P22): 5'-CTTATAAATTCGTAATGCGAAG-3'

SEQ ID No.10: (attB sequence of phage F80): 5'-AACACTTTCTTAAATGGTT-3'

SEQ ID No.11: (attP sequence of phage F80): 5'-AACACTTTCTTAAATTGTC-3'

SEQ ID No.12: (attB sequence of phage HP1): 5'-AAGGGATTTAAAATCCCTC-3'

SEQ ID No.13: (attP sequence of phage HP1): 5'-ATGGTATTTAAAATCCCTC-3'

SEQ ID No.14: (att sequence of plasmid pSAM2): 5'-TTCTCTGTCGGGGTGGCGGGATTTGAACCCACGACCTCTTCGTCCCGAA-3'

SEQ ID No.15: (Recognition sequence of the resolvase of the transposon Tn3): 5'-CGTCGAAATATTATAAATTATCAGACA-3'

SEQ ID No.16: oligonucleotide 6006: 5'-GATCTGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGACTGCAGATCT-3'

SEQ ID No.17: oligonucleotide 6008: 5'-GATCAGATCTGCAGTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCA-3'

SEQ ID No.18: (Sequence present in plasmid pXL2727 corresponding to the oligonucleotide 6008): 5'-GATCAGATCTGCAGTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCA-3'

SEQ ID No.19: (oligonucleotide 116418): 5'-CTTCTTCTTCTTCTTCTTCT-3'

SEQ ID No.20: (oligonucleotide 5542): 5'-AGCTTCTCGAGCTGCAGGATATCGAATTCGGATCCTCTAGAGCGGCCGCGAGCTCC-3'

SEQ ID No.21: (oligonucleotide 5543): 5'-AGCTGGAGCTCGCGGCCGCTCTAGAGGATCCGAATTCGATATCCTGCAGCTCGAGA-3'

SEQ ID No.22: sense oligonucleotide 6194: 5'-ACTAGTGGCCATGCATCCGCTCAAGTTAGTATAAAAAAGCAGGCTTCAG-3'

SEQ ID No.23: antisense oligonucleotide 6195: 5'-AGCTCTGAAGCCTGCTTTTTTATACTAACTTGAGCGGATGCATGGCCACTAGTAGCT-3'

SEQ ID No.24: sense oligonucleotide 6190: 5'-GCGTCTAGAACAGTATCGTGATGACAGAG-3'

SEQ ID No.25: antisense oligonucleotide 6191: 5'-GCCAAGCTTAGCTTTGCACTGGATTGCGA-3'

Bibliographic References

Ausubel et al. Current protocols in molecular biology 1987–1988. John Willey and Sons, New York.

Behr J. P. 1993. Acc. Chem. Res. 26:274–278.

Casadaban et al. 1983. Methods Enzymol. 100, 293–308.

Cotten et al. E. 1993. Curr. Biol. 4:705–710.

Giles, J. W. 1985. Am. Biotechnol., November/December.

Hasan et al. 1987. Gene 56:145–151.

Jain, R. K. 1987. Cancer Res. 47:3039–3051.

Landford et al. 1986. Cell 46:575–582.

Landy, A. 1989. Ann. Rev. Biochem. 58:913–949.

Maniatis, T., E. F. Fritsch, and J. Sambrook. 1989. Molecular cloning: a laboratory manual, second edition. Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, New York.

Nabel et al. 1992. Human Gene Therapy 3:399–410.

Podhajska et al. 1985. Gene 40:163:168.

Sadler et al. 1980. Gene, 8:279–300.

Sinha et al. 1984. Acids Res., 12, 4539–4557.

Stark et al. 1992. Trends Genet. 8:432–439.

Viera et al. 1982. Gene, 19, 259–268.

Wils et al. Biochem. Pharmacol. 48:1528–1530.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(iii) NUMBER OF SEQUENCES: 25

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

-continued

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

AATTGTGAAG CCTGCTTTTT TATACTAACT TGAGCGG 37

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

AATTCCGCTC AAGTTAGTAT AAAAAAGCAG GCTTCAC 37

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 57 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GATCCGAAGA AGAAGAAGAA GAAGAAGAAG AAGAAGAAGA AGAAGAAGAA GAAGAAC 57

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 57 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GATCGTTCTT CTTCTTCTTC TTCTTCTTCT TCTTCTTCTT CTTCTTCTTC TTCTTCG 57

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GAGGCTTCTT CTTCTTCTTC TTCTT 25

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

-continued

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CTGCTTTTTT ATACTAACTT G 21

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CAGCTTTTTT ATACTAAGTT G 21

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CAGCGCATTC GTAATGCGAA G 21

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CTTATAATTC GTAATGCGAA G 21

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

AACACTTTCT TAAATGGTT 19

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

-continued

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

AACACTTTCT TAAATTGTC 19

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

AAGGGATTTA AAATCCCTC 19

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

ATGGTATTTA AAATCCCTC 19

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 49 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

TTCTCTGTCG GGGTGGCGGG ATTTGAACCC ACGACCTCTT CGTCCCGAA 49

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CGTCGAAATA TTATAAATTA TCAGACA 27

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 66 base pairs
(B) TYPE: nucleic acid

-continued

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GATCTGAAGA AGAAGAAGAA GAAGAAGAAG AAGAAGAAGA AGAAGAAGAA GAAGAACTGC 60

AGATCT 66

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 66 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GATCAGATCT GCAGTTCTTC TTCTTCTTCT TCTTCTTCTT CTTCTTCTTC TTCTTCTTCT 60

TCTTCA 66

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 57 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GATCAGATCT GCAGTCTCTT CTTCTTCTTC TTCTTCTTCT TCTTCTTCTC TTCTTCA 57

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CTTCTTCTTC TTCTTCTTCT T 21

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 56 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

AGCTTCTCGA GCTGCAGGAT ATCGAATTCG GATCCTCTAG AGCGGCCGCG AGCTCC 56

-continued

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 56 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

AGCTGGAGCT CGCGGCCGCT CTAGAGGATC CGAATTCGAT ATCCTGCAGC TCGAGA 56

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 49 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

ACTAGTGGCC ATGCATCCGC TCAAGTTAGT ATAAAAAAGC AGGCTTCAG 49

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 57 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

AGCTCTGAAG CCTGCTTTTT TATACTAACT TGAGCGGATG CATGGCCACT AGTAGCT 57

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GCGTCTAGAA CAGTATCGTG ATGACAGAG 29

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GCCAAGCTTA GCTTTGCACT GGATTGCGA 29

What is claimed is:

1. A double-stranded DNA molecule, comprising an expression cassette containing a gene of interest under control of a transcription promoter and a transcription terminator active in a mammalian cell, wherein said molecule:

is in circular and supercoiled form,

lacks an origin of replication,

lacks a marker gene, and

comprises a region resulting from site-specific recombination between two sequences, said region being located outside the expression cassette.

2. The molecule according to claim 1, further comprising a sequence which interacts specifically with an oligonucleotide to form a triple helix by hybridization.

3. The molecule according to claim 2, wherein the sequence which forms a triple helix comprises from 5 to 30 base pairs.

4. The molecule according to claim 2, wherein the sequence which forms a triple helix is a homopurine-homopyrimidine sequence.

5. The molecule according to claim 1, wherein said region results from site-specific recombination between two att attachment sequences, two recognition sequences of a resolvase of a transposon, or two mrs sequences of plasmid RK2.

6. The molecule according to claim 1, further comprising an mrs sequence originating from a par locus of RK2.

7. The molecule according to claim 1, wherein the gene of interest is a nucleic acid coding for a therapeutic, vaccine, agricultural, or veterinary product.

8. The molecule according to claim 1, wherein said molecule is obtained by excision from a plasmid or chromosome by site-specific recombination.

9. A recombinant DNA comprising a polynucleotide comprising an expression cassette positioned between two sequences positioned in direct orientation, which recombine by site-specific recombination in the presence of a recombinase, wherein said expression cassette comprises a gene of interest under control of a transcription promoter and a transcription terminator active in a mammalian cell, and wherein said polynucleotide lacks an origin of replication and a marker gene.

10. The recombinant DNA according to claim 9 further comprising an origin of replication and, optionally, a marker gene, wherein the origin of replication and optional marker gene are located outside said polynucleotide.

11. The recombinant DNA according to claim 9, wherein the recombinase is a recombinase of an integrase family of phage lambda or of a resolvase family of transposon Tn3.

12. The recombinant DNA according to claim 9, wherein the two sequences which recombine by site-specific recombination are derived from a bacteriophage.

13. The recombinant DNA according to claim 12, wherein the two sequences which recombine by site-specific recombination consist of att attachment sequences of a bacteriophage or sequences derived therefrom.

14. The recombinant DNA according to claim 13, wherein the two sequences which recombine by site-specific recombination consist of attachment sequences of bacteriophage lambda, P22, Φ080, P1, or HP1, of plasmid pSAM2, or of sequences derived therefrom.

15. The recombinant DNA according to claim 14, wherein the sequences which recombine by site-specific recombination comprise all or part of SEQ ID Nos. 1, 2, 6, 7, 8, 9, 10, 11, 12, 13, or 14.

16. The recombinant DNA according to claim 12, wherein the two sequences which recombine by site-specific recombination are derived from bacteriophage P1.

17. A plasmid comprising:

a) a bacterial origin of replication and optionally, a marker gene; and

b) a polynucleotide comprising an expression cassette positioned between attP and attB sequences of a bacteriophage lambda, P22, Φ80, P1, or HP1, or of plasmid pSAM2, positioned in direct orientation, which recombine by site-specific recombination in the presence of a recombinase, wherein said expression cassette comprises a gene of interest under control of a transcription promoter and a transcription terminator active in a mammalian cell, and wherein said polynucleotide lacks an origin of replication and a marker gene.

18. The plasmid according to claim 17, wherein the attP and attB sequences which recombine by site-specific recombination are attachment sequences of bacteriophage lambda.

19. A plasmid comprising:

a) a bacterial origin of replication and optionally, a marker gene; and

b) a polynucleotide comprising an expression cassette positioned between two inverted repeat sequences of bacteriophage P1 (loxP region) positioned in direct orientation, which recombine by site specific recombination in the presence of a recombinase; wherein the expression cassette comprises a gene of interest under control of a transcription promoter and a transcription terminator active in a mammalian cell, and wherein the polynucleotide lacks an origin of replication and a marker gene.

20. The recombinant DNA according to claim 9, wherein the two sequences which recombine by site-specific recombination are derived from a transposon.

21. The recombinant DNA according to claim 20, wherein the two sequences which recombine by site-specific recombination consist of recognition sequences of a resolvase of a transposon Tn3, Tn21, or Tn522, or sequences derived therefrom.

22. The recombinant DNA according to claim 21, wherein the two sequences which recombine by site-specific recombination comprise all or part of sequence SEQ ID No.15.

23. The recombinant DNA according to claim 9, wherein the two sequences which recombine by site-specific recombination are derived from a par region of plasmid RP4.

24. The recombinant DNA according to claim 9, wherein the expression cassette further comprises a sequence which interacts specifically with an oligonucleotide to form a triple helix by hybridization.

25. A plasmid comprising:

a) an origin of replication and optionally, a marker gene; and

b) a polynucleotide comprising at least one gene of interest and a sequence which interacts specifically with an oligonucleotide to form a triple helix by hybridization, wherein the at least one gene of interest and the oligonucleotide interacting sequence are positioned between two sequences positioned in direct orientation, which recombine by site-specific recombination in the presence of a recombinase, and wherein the polynucleotide lacks an origin of replication and a marker gene.

26. A plasmid comprising:

a) an origin of replication and optionally, a marker gene; and

b) a polynucleotide comprising at least one gene of interest and an mrs sequence originating from a par

locus of plasmid RK2, wherein the at least one gene of interest and the mrs sequence are positioned between two sequences positioned in direct orientation, which recombine by site-specific recombination in the presence of a recombinase, and wherein the polynucleotide lacks an origin of replication and a marker gene.

27. The plasmid according to claim **26**, wherein the polynucleotide further comprises a sequence which interacts specifically with an oligonucleotide to form a triple helix by hybridization, wherein the oligonucleotide interacting sequence is placed between the two sequences positioned in direct orientation, which recombine by site-specific recombination.

28. A plasmid comprising:

a) an origin of replication and optionally, a marker gene; and

b) a polynucleotide comprising:

1) a first set of two sequences positioned in direct orientation, which recombine by integrase-dependent site-specific recombination;

2) a second set of two sequences positioned in direct orientation, which recombine by resolvase-dependent site-specific recombination;

3) at least one gene of interest; and,

4) optionally, a sequence which interacts specifically with an oligonucleotide to form a triple helix by hybridization,

wherein each integrase-dependent sequence of 1) is positioned next to a resolvase-dependent sequence of 2) and wherein the at least one gene of 3) and the optional oligonucleotide interacting sequence of 4) are placed between the integrase-dependent/resolvase-dependent sequences, and wherein the polynucleotide lacks an origin of replication and a marker gene.

29. A cultured recombinant cell comprising one or more copies of the recombinant DNA according to claim **9** inserted into its genome.

30. A cultured recombinant cell comprising the recombinant DNA according to claim **10**.

31. The cultured recombinant cell according to claim **30**, wherein said cell is a bacterium.

32. The cultured recombinant cell according to claim **30**, wherein said cell is a eukaryotic cell.

33. The cultured recombinant cell according to claim **31**, wherein the bacterium is *Escherichia coli* D1210HP with accession number I-2314.

34. A method for preparation of the DNA molecule according to claim 1, comprising culturing 1) a host cell comprising a recombinant DNA comprising a nucleic acid consisting of an expression cassette positioned between two sequences positioned in direct orientation, which recombine by site-specific recombination in the presence of a recombinase, and wherein the expression cassette comprises a gene of interest under control of a transcription promoter and a transcription terminator active in a mammalian cell with 2) a recombinase, whereby site-specific recombination occurs between the two sequences positioned in direct orientation.

35. The method according to claim **34**, wherein said expression cassette is positioned between two bacteriophage sequences, which are positioned in direct orientation and recombine by site-specific recombination.

36. The method according to claim **34**, wherein the cultured host cell is brought into contact with the recombinase by transfecting or infecting the cultured host cell with a plasmid or a phage containing a gene for the recombinase.

37. The method according to claim **34**, wherein the cultured host cell is brought into contact with the recombinase by inducing expression of a gene coding for the recombinase, wherein the gene is present in the host cell.

38. The method according to claim **37**, wherein the host cell comprises within its genome a recombinase gene having temperature-regulated expression, and wherein the cultured host cell is brought into contact with the recombinase by culturing the host cell at an induction temperature of the recombinase gene, whereby expression of the recombinase gene is induced.

39. The method according to claim **38**, wherein the host cell comprises a lysogenic phage integrated in its genome and wherein the lysogenic phage comprises the gene for the recombinase.

40. A method for preparation of the DNA molecule according to claim 1, comprising combining:

a) a replicative plasmid comprising:

1) an origin of replication and optionally, a marker gene; and

2) a polynucleotide comprising an expression cassette positioned between two sequences positioned in direct orientation, which recombine by site-specific recombination in the presence of a recombinase, wherein the expression cassette comprises a gene of interest under control of a transcription promoter and a transcription terminator active in a mammalian cell, and wherein the polynucleotide lacks an origin of replication and a marker gene; and

b) a recombinase, whereby site-specific recombination occurs between the two sequences of 2) positioned in direct orientation.

41. The method according to claim **34**, further comprising purifying a minicircle formed by said site-specific recombination.

42. The method according to claim **41**, wherein the minicircle is purified by contacting the minicircle with a specific oligonucleotide that is grafted onto a support, whereby a triple helix is formed by hybridization of said specific oligonucleotide with a specific sequence present in the minicircle.

43. The recombinant DNA according to claim 9, wherein the two sequences which recombine by site-specific recombination are from a 2μ plasmid.

* * * * *

EXHIBIT 2

Article No. jmbi.1999.2730 available online at http://www.idealibrary.com on IDEAL J. Mol. Biol. (1999) 288, 825–836

C1

JMB XP-002164454

Alterations in the Directionality of λ Site-specific Recombination Catalyzed by Mutant Integrases *in Vivo*

Nicole Christ and Peter Dröge*

Institute of Genetics, University of Cologne, Weyertal 121 D-50931, Cologne, Germany

**Corresponding author*

Phage λ integrative and excisive recombination normally proceeds by a pair of sequential strand exchanges. During the first exchange reaction, the "top" strand in each recombination site is cleaved, exchanged, and religated generating a Holliday junction intermediate. This intermediate DNA structure is resolved through a pair of reciprocal "bottom" strand exchanges, leading to recombinant products. The strict co-ordination of exchange reactions ensures religation between correct partner strands only. Here we show that the directionality of recombination is altered *in vivo* by two mutant integrases, Int-h (E174 K) and a double mutant Int-h/218 (E174 K/E218 K). This change in directionality leads to deletion instead of inversion on substrates that carry inverted attachment sites and, depending on the pair of target sites employed, requires the presence or absence of integration host factor. Neither Fis nor Xis is involved in deletion. Sequence analyses of deletion products reveal that the newly generated hybrid attachment site exhibits a reversed genetic polarity. We demonstrate that only one of two possible hybrid site configurations is generated and discuss two pathways leading to deletion. In the first, deletion results from a wrong alignment of the two recombination sites within the synaptic complex. In the second pathway, the uncoordinated cleavage by the mutant integrases of all four DNA strands present in a conventional Holliday junction intermediate leads to two double-stranded breaks, whereby the subsequent rejoining between "wrong" partner strands appears restricted to only two strands.

Keywords: site-specific recombination; λ integrase; strand exchange; integration host factor; Holliday junction isomerization

Introduction

The phage λ-encoded integrase protein (Int) is the prototype of the so-called integrase family which catalyzes conservative site-specific recombination between two DNA target sites. Int executes both the integration and excision of the phage into and out of the *Escherichia coli* genome, respectively. The structure of the catalytic domain of Int has recently been solved (Kwon *et al.*, 1997). The Int system represents, therefore, one of the best understood recombination systems (for reviews, see Landy, 1989, 1993; Sadowski, 1993; Nash, 1996; Hallet & Sherratt, 1997; Yang & Mizuuchi, 1997).

Integrative and excisive recombination occurs between pairs of attachment sites, termed *attP*/*attB* and *attL*/*attR*, respectively. Each *att* sequence is composed of two core Int binding sites separated by a seven base-pair overlap region. The overlap sequence is identical in all wild-type *att* sites, and identity is a prerequisite for efficient recombination. In addition to core sites where strand cleavage and religation occurs, each site except *attB* contains additional Int binding sites, so-called arm sites. A varying number of flanking recognition sequences for the accessory DNA-bending proteins integration host factor (IHF), factor for inversion stimulation (Fis), or the phage-encoded Xis protein are also present in the flanking regions, again with the exception of *attB*. Int is a heterobivalent DNA-binding protein and, with assistance from the accessory proteins, is able to bind simultaneously to core and arm sites within the same *att* site (Moitoso de Vargas *et al.*, 1988, 1989). Depending on both the presence and number of accessory factors,

Abbreviations used: Int, phage-encoded integrase; IHF, integration host factor; Fis, factor for inversion stimulation; Xis, phage-encoded excisionase.

E-mail address of the corresponding author: p.droege@uni-koeln.de

0022-2836/99/200825–12 $30.00/0

the resulting nucleoprotein structures at *attP*, *attL*, and *attR* exhibit different architectures, and it is this difference that controls for the directionality of the reaction, i.e. integration or excision (Moitoso de Vargas & Landy, 1991).

In the first step leading to integrative recombination, a specialized nucleoprotein structure, the intasome, is formed between Int, IHF, and supercoiled *attP* (Better *et al.*, 1982; Richet *et al.*, 1986). The second step involves the pairing of the intasome with protein-free *attB*, the latter consists only of two core sites and the overlap region. Hence, Int monomers which catalytically act upon *attB* are in this case exclusively provided from *attP* (Richet *et al.*, 1988; Patsey & Bruist, 1995). There is no DNA topological constraint imposed on synapsis between *attP* and *attB*, which explains why both direct and inverted pairs of *att* sites present on the same DNA molecule are efficiently recombined *in vitro* and *in vivo*, leading to deletion or inversion of the intervening DNA segment, respectively.

In the third step, Int catalyzes a reciprocal exchange of the "top" strands at the left boundary of the overlap region, which results in a Holliday junction recombination intermediate (Figure 1(a) and (b)). This intermediate DNA structure is resolved by exchange of the "bottom" strands at the right boundary of the overlap region (Figure 1(b) to (d)). Thus, Int executes an ordered, sequential pair of strand exchanges, i.e. cleavage, exchange, and rejoining of one pair of recombination partner strands is completed before initiating the same reactions on the other pair of strands (Nunes-Düby *et al.*, 1987; Kitts & Nash, 1988). During these reactions, Int becomes covalently attached *in cis* to the broken DNA strand through a 3′-phosphotyrosine linkage, which is subsequently resolved when the 5′-hydroxyl group of the invading strand attacks the linkage and displaces the recombinase (Figure 1(c) and (d)); Burgin & Nash, 1992; Nunes-Düby *et al.*, 1994). Neither supercoiling of *attP* nor the presence of IHF seems to be required for catalysis of these chemical reactions. Integrative recombination eventually leads to the formation of *attL* and *attR*, which are the targets for excision (Figure 1(d)).

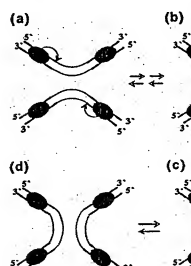


Figure 1. Conventional strand exchange during λ integrative recombination. (a) *attB* (genetic polarity BOB′) and *attP* (POP′) are aligned in antiparallel orientation. The Int monomer (filled oval) bound to either the B arm (marked B) or P arm (marked P) initiates a nucleophilic attack (filled arrows) against the top strand within each *att* site. (b) Reciprocal strand exchanges between top strands has been completed, leading to a Holliday junction intermediate structure. (c) The Int monomers bound to the B′ and P′ arm have cleaved the bottom strands at the right ends of both overlap regions and are thus covalently linked to the DNA. (d) Reciprocal bottom strand exchange has been completed leading to *attL* (genetic polarity BOP′) and *attR* (POB′), which are the natural targets for excisive recombination.

Excisive recombination is genetically the exact reversal of integration, but employs different nucleoprotein structures on *attL*/*attR*. In addition to IHF, which again serves mainly as an architectural protein at these sites (Goodman *et al.*, 1992), the λ phage-encoded Xis protein is required for the formation of a recombinogenic nucleoprotein complex at *attR*. Thus, in contrast to integrative recombination, two separate nucleoprotein structures are formed before synapsis occurs by random collision (Kim & Landy, 1992). It was also shown that Int, in the absence of any accessory proteins, can align *attP* and *attL* within a bi-molecular complex, and can even recombine pairs of *attL* (Segall & Nash, 1993, 1996). However, the order of strand exchange during excisive recombination is the same as observed for integrative recombination.

According to a recently proposed model, the switch from top to bottom strand exchange involves isomerization of the Holliday junction with concomitant restacking of base-pairs within the overlap region (Nunes-Düby *et al.*, 1995; Azaro & Landy, 1997). How Int controls this step and, thus, ensures the order of strand exchange during integrative and excisive recombination is not known. There is evidence that specific protein-protein interactions are required between at least three Int protomers bound to core sites within a Holliday junction (Franz & Landy, 1990; Kho & Landy, 1994). In addition, Int mutants have been isolated which indicate that specific interactions between Int monomers are important in the co-ordination of strand cleavage events within this intermediate structure (Han *et al.*, 1994). Conformational changes of both the catalytic domain and the molecular interface within and between Int monomers, respectively, are probably involved in co-ordinating the sequence of strand exchanges. This can be inferred from the structure of the Cre-*lox*P Holliday junction intermediate (Gopaul *et al.*, 1998). The exact roles for IHF and Xis in this co-ordination is also unclear. There is evidence that both proteins control the efficiency of resolution of Holliday

junctions in the direction of recombinant products (Franz & Landy, 1995).

Here, we show that two mutant Int variants, E174 K and E174 K/E218 K, alter the directionality of recombination reactions *in vivo*, leading to deletion instead of inversion on substrates that carry two *att* sites as inverted repeats. The efficiency of this reaction depends on the type of *att* sites employed and the presence or absence of IHF. Nearly 100% deletion occurs, for example, with substrates bearing inverted *att*L and *att*P sites in the presence of IHF. However, neither Xis nor Fis is involved in deletion. Based on DNA sequence information obtained from various deletion products, we discuss two possible mechanisms leading to deletion.

Results

Integrase mutants

Here, we analyze the *in vivo* catalytic activities of two Int mutants. The first one, termed Int-h, was originally identified by Miller *et al.* (1980) in a screen for λ mutants that overcome the block imposed on recombination by the *him*A42 mutation of *E. coli*. The mutation replaces a glutamate residue with a lysine residue at codon 174 and maps within the N-terminal region of the catalytic domain (Figure 2). Subsequent purification and characterization of Int-h revealed that it promotes integrative and excisive recombination in the absence of accessory proteins and supercoiling, albeit with significantly reduced efficiencies (Lange-Gustafson & Nash, 1984). It is proposed that this mutation results in an enhanced affinity for core sites, which would account for the increased frequency of *in vivo* and *in vitro* integration into secondary *att* sites that deviate from the wild-type *att*B sequence (Miller *et al.*, 1980; Patsey & Bruist, 1995).

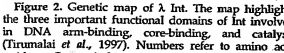


Figure 2. Genetic map of λ Int. The map highlights the three important functional domains of Int involved in DNA arm-binding, core-binding, and catalysis (Tirumalai *et al.*, 1997). Numbers refer to amino acid residues. Open circles indicate the position of Ala125 and Ala126, which make close contact with bases in the core binding site (Tirumalai *et al.*, 1998). Filled circles mark the positions of three strictly conserved residues within the Int family of recombinases, i.e. Arg212, His308, and Arg311 (reviewed by Nunes-Düby *et al.*, 1998). The asterisk marks the active site residue, tyrosine 342. Crosses demarcate the positions of four residues that seem involved in determining recombination specificity with respect to target sites (Yagil *et al.*, 1995; Dorgai *et al.*, 1995). Int-h and Int-h/218 mark the corresponding amino acid changes within two mutant Int variants analyzed in the present study (see the text).

Our initial goal was to further enhance the ability of Int-h to perform integration into secondary *att* sites in the absence of supercoiling. We thus introduced, by PCR-directed mutagenesis, a second mutation into the catalytic domain of Int-h, which replaces a glutamate residue at codon 218 with a lysine residue. Our choice for mutating this specific residue is based on a recent finding (Wu *et al.*, 1997) that lysine at this position improves binding of wild-type Int to core sites, presumably through non-specific contact(s) with the DNA backbone. We will subsequently refer to the double mutant as Int-h/218 (Figure 2).

In vivo catalytic activities of Int-h and Int-h/218

We first tested whether Int-h/218 retains the ability of Int-h to promote integrative recombination in the absence of IHF *in vivo*. Expression vectors for Int, Int-h, and Int-h/218 were co-transformed with pNC1, which carries *att*B and *att*P as inverted repeat (Table 1), into either *E. coli* strain *CSH26* or an isogenic variant, *CSH* 26AIHF. In the latter strain, both genes encoding for subunits of IHF have been destroyed by transposition. Resulting single colonies were cultivated overnight, plasmid DNA isolated, subjected to restriction digests, and analyzed by agarose gel electrophoresis.

Int and the two variants perform efficient integrative recombination in the presence of IHF, leading to 100% inversion without induction of gene expression by IPTG (Figure 3, lanes 4 to 7). Hence, leaky Int expression from the P_{trc} promoter in the presence of *lac* repressor bound to its operator sequences is sufficient to promote complete inversion over the time-course of the experiment. However, in the absence of IHF, Int is completely inactive and inversion catalyzed by Int-h is barely detectable (lanes 8 and 9). Int-h/218 executes inversion with comparable low efficiency, but we were surprised to detect a more prominent product band migrating above that of the expression vector (demarcated del.; Figure 3, lane 10). The yield of this new product increases over time (lane 11), but varies considerably between experiments (data not shown). The same product could also be detected with Int-h in some but not all experiments (Table 1).

Undigested plasmid DNA derived from a sample obtained with Int-h/218 was resolved through electrophoresis, the new product DNA isolated and re-transformed into *E. coli*. Restriction analysis of plasmid DNAs from single colonies revealed that the products contain a deletion between *att*B and *att*P (data not shown). DNA sequencing confirmed that notion. We found two different products. The first, termed *att*Δ1, results from recombination that joins the left core site of *att*B, the so-called B arm, to the left core site of *att*P, the P arm (see Figure 6(a)). In the following, we refer to this type of new *att* site with reversed genetic polarity, i.e. BOP instead of BOP′, as a

Table 1. DNA substrates and catalytic activities of mutant and wild-type integrase

substrate	Int-h and Int-h/218 inversion +IHF	Int-h and Int-h/218 inversion -IHF	Int-h and Int-h/218 deletion +IHF	Int-h and Int-h/218 deletion -IHF	wild-type Int inversion +IHF	wild-type Int inversion -IHF	wild-type Int deletion +IHF	wild-type Int deletion -IHF
pNC1	+[a]	+	-	+[b]	+	-	-	-
pNC2	-[c]	+	-	+	-	-	-	-
pNC3	-	+	-	-	-	-	-	-
pNC4	+	+	+	-	+	+	-	-
pNC5	-[d]	+	-[d]	-	+	+	-	-
pNC6	+[b]	-	+	-	+	-	-	-
pNC7	-	-	+[e]	-	-	-	-	-
pNC8	-	-	-	-	-	-	-	-
pNC9	-	-	+	+[b]	-	-	+	-
pNC10	-	-	+	+[b]	-	-	+	-
pNC11	-	-	+	+	-	-	+	+
pNC12	-	nd[f]	+[b]	nd	nd	nd	nd	nd

Schematic representation of recombination substrates. The relative positions and orientations of relevant genetic elements present on DNA substrates employed in this study are indicated. Arrows demarcate the attachment (*att*) sites, the open rectangle marks the position of the kanamycin resistance gene, and the filled rectangle represents the pACYC origin of replication. *attP** refers to the presence of a nucleotide change within the overlap region (see the text). The arm sites within each att and relevant positions of cleavage sites for restriction enzymes are also indicated.

[a] The (+) sign refers to the presence of the corresponding DNA band after ethidium staining.
[b] Int-h is significantly less active than Int-h 218.
[c] Inversion observed after 48 hours
[d] Complete loss of substrate vector.
[e] Barely detectable with Int-h/218 and Int-h after 24 and 48 hours, respectively.
[f] Not determined. The expected recombination products, i.e. inversion for pNC1 through pNC8 and deletion for pNC9 through pNC12, are boxed.

hybrid site. Our analysis also revealed that in the particular orientation chosen to depict *attΔ*1 (BOP) with respect to *attB*/*attP* (see Figure 6(a)), the top strand of the overlap region within *attΔ*1 is derived from the bottom strand of *attP*. However, the top strand of the overlap in the second product, *attΔ*2, is derived from the top strand of *attB* (see Figure 6(a)). From a total of 12 analyzed sequences, each derived from a single colony after re-transformation, we found that eight contain the overlap provided from *attP* while four carry the overlap region from *attB*.

Int-h is known to recombine pairs of *att* sites which differ in their overlap sequence by one or more base-pairs more efficiently than wild-type Int (Kitts & Nash, 1987; Patsey & Bruist, 1995). This feature is, at least in part, ascribed to the enhanced affinity of Int-h to core sites. In order to test whether deletion can be detected and, if so, how Int-h/218 performs on such a substrate, we constructed pNC6, which contains *att*B and a variant form of *attP* (*attP**) as inverted repeats (Table 1). Within *attP** a guanine base replaces the third nucleotide, a thymine residue, of the overlap sequence. pNC6 was co-transformed with Int expression vectors, and plasmid DNA was subjected to restriction analysis. While in the absence of IHF, all three Int variants are completely inactive in integrative recombination (data not shown; Table 1), both Int-h and, with enhanced efficiency, Int-h/218 execute inversion in the presence of IHF (Figure 4, lanes 2 and 3). However, both mutants in addition generate a second prominent product migrating at the top of the gel. The new product was isolated and amplified as before, and its sequence revealed that the B arm of *attB* is joined to the P arm of *attP*, as observed before with pNC1. In this case, however, we found that the

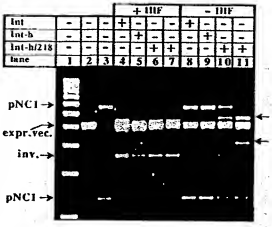


Figure 3. *In vivo* catalytic activities of wild-type and mutant Int on pNC1. pNC1, which carries *attB* and *attP* as inverted repeats, was co-transformed into *E. coli* with expression vectors for either wild-type or mutant Int (see Materials and Methods). Isolated plasmid DNA was incubated with *AvaI* and the resulting restriction fragments analyzed through agarose gel electrophoresis. We show a gel after ethidium bromide staining. Lane 1, kb marker ladder; lane 2, expression vector alone; lane 3, unrecombined pNC1 (note that the third restriction fragment is not shown); lanes 4 to 6, DNA isolated from cells expressing Int, Int-h, or Int-h/218, respectively, in the presence of IHF; lane 7, same as lane 6, but DNA was obtained from a different colony; lanes 8 to 10, DNA isolated from CSH26ΔIHF cells expressing Int, Int-h, or Int-h/218, respectively; lane 11, same as lane 10, but DNA isolated after an additional 24 hours incubation. pNC1, unrecombined substrate DNA; expr. vec., Int expression vector; inv., one of two product bands that results from inversion (the second co-migrates with the expression vector); del., deletion product which is cleaved only once due to the absence of a second restriction site. The asterisk marks the position of a third product band which is present in some experiments and which results from homologous recombination as determined by restriction analysis.

overlap originates in 11 out of 12 analyzed sequences from *attB* (data not shown).

Intrigued by the relatively high yield of deletion products obtained with pNC6, we investigated whether other pairs of inverted *att* sites can lead to deletion. We therefore constructed two substrates for excisive recombination, pNC2 and pNC7, which contain *attL* and *attR* as inverted repeats (Table 1). The *attR* site in pNC7, termed *attR**, carries the same mutated overlap sequence as present in *attP**. We found that in the absence of IHF (and Xis), Int-h and Int-h/218 deleted the segment between *att* sites in pNC2 with the same efficiency as observed with pNC1. Deletion on pNC7 was barely detectable with Int-h/218 in the presence of IHF, and could only be detected after 48 hours of *in vivo* incubation with Int-h. However, it is noteworthy that we were unable to detect inversion on pNC7 either in the presence or absence of IHF (Table 1). This may indicate that deletion requires initial steps of the conventional strand exchange pathway (see Discussion).

The wild-type *att* sites tested so far are converted by conventional strand exchange into the expected recombination products (Table 1; see Figure 6(a)). It is impossible to determine from these experiments whether one or both pairs of sites, i.e. (*attB*/*attP*) or/and (*attL*/*attR*), are the substrates for deletion. We therefore constructed a substrate that cannot be altered by conventional strand exchange, pNC4, which carries *attP* and *attL* as inverted repeats (Table 1). If recombination occurs between these sites, this will lead to inversion. Their genetic polarities do not change, however (see Figure 6(b)).

We found that in the presence of IHF, nearly 100% of pNC4 is converted into a deletion product upon reaction with either Int-h or Int-h/218 (Figure 5, lanes 5 and 6). Wild-type Int executes only inversion (lane 4). In the absence of IHF, Int and the two variants exclusively catalyze inversion (lanes 7 to 9). Samples of undigested deletion products obtained from Int-h/218 and Int-h were isolated, re-transformed, and DNA from ten and nine colonies, respectively, subjected to DNA sequencing. In each case, we found two hybrid *att* sites. The first, termed *attΔ3*, was present three or two times, respectively, and contains the B arm joined to the P arm (Figure 6(b)). In the orientation chosen for *attΔ3* in Figure 6(b) (BOP), both arm sites are separated by the overlap provided from the bottom strand of *attP*. In the second product, *attΔ4*, which was found seven times in each case, the overlap originates from the top strand of *attL*. From this analysis, we conclude that inverted *attL* and *attP* can serve as substrate for a highly efficient change in the directionality of recombination by mutant Int in the presence of IHF, leading to almost 100% deletion instead of inversion.

We further tested whether other *E. coli* proteins, in addition to IHF, may play a role in deletion. At the present stage using various *E. coli* strains (see Materials and Methods), we found that *recA*, *recB*, and *recC* are not required for deletion with pNC4

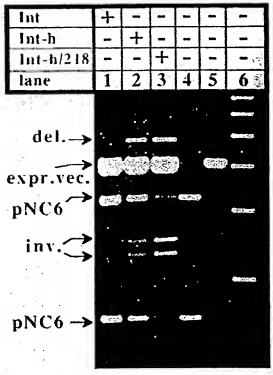


Figure 4. *In vivo* catalytic activities of wild-type and mutant Int on pNC6. pNC6, which carries wild-type *att*B and a mutated *att*P (*att*P*) as inverted repeats was co-transformed with Int expression vectors as before, and plasmid DNA isolated and subjected to gel electrophoresis after restriction digest with *Asu*I (Table 1). Lanes 1 to 3, DNA isolated from cells expressing Int, Int-h, or Int-h/218, respectively, in the presence of IHF; lane 4, unrecombined pNC6; lane 5, expression vector alone; lane 6, kb marker ladder. inv. demarcates the position of the two product bands that result from inversion, del. points to the position of the linearized deletion product.

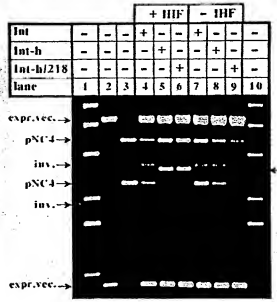


Figure 5. *In vivo* catalytic activities of wild-type and mutant Int on pNC4. pNC4, which carries inverted *att*L and *att*P, was co-transformed with Int expression vectors and isolated plasmid DNA processed as described before, except that *Bam*HI was used as endonuclease (compare to Table 1). Lanes 1 and 10, kb marker ladder; lane 2, expression vector alone; lane 3, unrecombined pNC4; lanes 4 to 6, DNA isolated from cells expressing Int, Int-h, or Int-h/218, respectively, in the presence of IHF; lanes 7 to 9, same as lanes 4 to 6, but in the absence of IHF. pNC4 demarcates the position of the three DNA fragments that result from digestion of unrecombined pNC4 (note that the two smaller fragments exhibit the same length and, thus, co-migrate); inv. indicates the position of two of three fragments that result from inversion (the third fragment exhibits the same size as the largest fragment obtained from unrecombined DNA); del. marks the position of one of two DNA fragments that result from deletion (the size of the second fragment does not change as a result of deletion).

and pNC6. In adition, deletion occurs in the absence of Fis on a derivative of pNC6 (data not shown).

Only one of two possible new hybrid *att* sites is generated

So far we have analyzed the products that result from deletion on substrates that contain inverted *att* sites, leading to the identification of a new hybrid site with reversed genetic polarity, i.e. BOP instead of BOP'. In order to address the question whether mutant Int can also generate the corresponding second hybrid site composed of B' (or P') arm, overlap, and P' arm (B'OP'), we first constructed a series of substrates that contain different pairs of *att* sites as direct repeats (pNC9 to pNC12; Table 1). If the mutant Int proteins execute a complete set of alternative strand exchange reactions which involves all four DNA strands present within a synaptic complex, this should result in inversion instead of deletion. However, we found that deletion occurs on all four substrates, but inversion between directly repeated *att* sites was not detectable, in either the presence or absence of IHF. These experiments also revealed that Int-h/218 is significantly more active than Int-h in catalyzing deletion on pNC9 and pNC10 in the absence of IHF (Table 1). Hence, Int-h/218 exhibits an enhanced ability to execute recombination on wild-type *att* sites in the absence of accessory factors IHF and Xis.

Our failure to detect the second hybrid site could so far be due to the possibility that a functional synaptic complex cannot be formed because an unknown topological constraint is imposed on synapsis with substrates carrying *att* sites as direct repeats. In order to exclude this possibility, we again employed the pair of *att*L and *att*P which

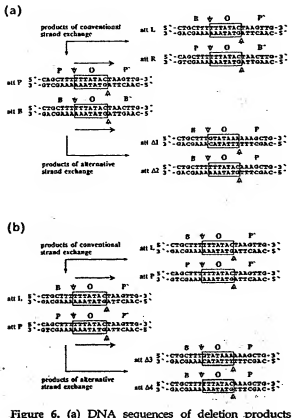


Figure 6. (a) DNA sequences of deletion products obtained with pNC1. The sequence of the core and overlap regions from *att*B and *att*P are shown at the left side. Both *att* sites are aligned in parallel, as indicated by black arrows. The capital letters B, B', P, and P' mark the corresponding core site sequences. The two overlap sequences are boxed and marked as O. The open arrow heads point to the positions of top and bottom strand cleavage by Int. Depicted at the top right are the sequences and genetic polarities of the two products, *att*L and *att*R, that result from conventional integrative recombination. Shown at the bottom right are the DNA sequences and genetic polarities of the two hybrid *att* sites, termed *att*Δ1 and *att*Δ2, which are present on deletion products. Note that in *att*Δ1 the overlap sequence is provided from the P arm, while that present in *att*Δ2 comes from the B arm. (b) DNA sequences of deletion products obtained with pNC4. The sequences of *att*L and *att*P are depicted at the left side. Both sites are aligned again in parallel. Symbols are as described in (a). The two products of conventional strand exchange are shown at the top right. Note that the composition of sites with respect to both the core sites and their flanking regions does not change during conventional strand exchange. Depicted at the bottom right are the sequences of the two hybrid *att* sites, *att*Δ3 and *att*Δ4, which are identical with *att*Δ1 and *att*Δ2, respectively.

gives the highest yield of deletion products with pNC4. In this case, however, we placed the inverted *att* sites in a different orientation with respect to both the plasmid origin and the resistance marker gene (pNC5; Table 1). If deletion occurs, a second hybrid *att* site (P'OP') should be generated that can be propagated by plasmid replication. The first identified site, BOP, will in this case be lost because the deleted DNA segment does not contain a replication origin. While we were able to detect inversion in the absence of IHF, products that result from deletion in the presence of IHF are missing. Instead, we observed that cells completely degrade pNC5 when either Int-h or Int-h/218 is present (Table 1). To test whether plasmid degradation is due to the instability of the expected recombination product carrying P'OP', we constructed a plasmid designated pP'OP' which contains the equivalent sequence of one of the two expected recombination products (see Materials and Methods). We found, however, that pP'OP' is stably propagated in the presence of Int-h (data not shown). Hence, the instability of pNC5 cannot be traced back to the instability of the expected recombination product carrying the second hybrid site. These results in conjunction with our failure to detect inversion on substrates pNC9 to pNC12 might indicate that both mutant Int proteins efficiently execute strand cleavage without subsequent ligation to generate the second hybrid *att* site.

Discussion

Integrative and excisive recombination performed by Int normally lead to inversion of DNA segments when the corresponding pair of target sites is present as an inverted repeat on the same DNA molecule. We have demonstrated in this study that mutant Int, in addition, executes an alternative reaction which leads to deletion. The most efficient reaction, resulting in nearly 100% deletion in the presence of IHF, occurs on pNC4 which carries *att*L and *att*P (Figure 5).

One possible model to account for the results is that the two mutant Int proteins have lost the ability to distinguish between the core binding sites present to the left and to the right of the overlap region. This would allow synapsis between two *att* sites in the wrong orientation. Reciprocal top strand exchange would then lead to mispaired top strands because of non-complementary bottom strands present in the resulting Holliday junction. Despite these heterologies, catalytic events may proceed normally due to the presence of mutant Int proteins with an enhanced affinity for core sites, and the Holliday junction may eventually be resolved through reciprocal bottom strand exchanges. The resulting heteroduplex structures at the overlap region of both hybrid *att* sites could then be resolved through repair and/or plasmid replication.

While this presumably represents the most simple scenario leading to deletion, we think it is unlikely for two reasons. First, the model predicts that two hybrid *att* sites should be generated due to reciprocal top and bottom strand exchanges. This should lead to inversion on substrates pNC9 to pNC12, and to deletion on pNC4 and pNC5.

Our results show, however, that inversion is not detectable, that deletion occurs only on pNC4 and not on pNC5, and that pNC5 is lost despite selection for the antibiotic resistance on this plasmid (Table 1). Second, the model predicts that pre-existing heterologies between overlap sequences of a pair of *att* sites should not be important because the strands will eventually be mispaired in a Holliday junction after the first strand exchange is completed. Hence, deletion on pNC1 and pNC6, the latter carries *attP**, should occur under the same conditions, i.e. in the presence or absence of IHF, and presumably with comparable efficiencies. The results show, however, that deletion on pNC1 occurs only in the absence of IHF while deletion on pNC6 is observed only in the presence of IHF and, in addition, occurs with an enhanced efficiency (Figures 3 and 4). Control experiments show that recombination between *attP** and a variant of *attB* carrying the same nucleotide exchange within the overlap region proceeds normally in the presence of IHF (data not shown). It is unlikely, therefore, that the different IHF requirements for deletion on pNC1 and pNC6 are due to a sequence effect imposed by the overlap in *attP**.

In the following, we propose a second possible pathway leading to deletion. We will focus our discussion primarily on deletion observed with pNC4 because this particular substrate has the advantage that the composition of *attL* and *attP* does not change during the course of conventional strand exchange (Figure 6(b)). It is reasonable to assume that in the first step, nucleoprotein structures separately assemble on *attL* and *attP* in the presence of IHF (Segall & Nash, 1993). After site synapsis in the correct orientation, the first strand exchange is completed (Figure 7(a) and (b)). This results in a Holliday junction, which can be resolved normally by reciprocal "bottom" strand exchanges (compare to Figure 1). However, we think that during a subsequent isomerization step, which is required to switch from top to bottom strand exchange (Nunes-Düby *et al.*, 1995; Azaro & Landy, 1997), Int-h and Int-h/218 accidentally cleave all four strands either sequentially or simultaneously. This will lead to two double-stranded breaks (Figure 7(c)). Based on DNA sequence information obtained from deletion products, we conclude that the 5′-OH group from the overlap strand extending the P arm (labelled p) engages in a nucleophilic attack on the 3′-phosphotyrosine linkage between Int-h and the B arm, hence replacing the recombinase. Likewise, the overlap strand still linked to the B arm (labelled b) attacks the Int-DNA linkage at the P arm (Figure 7(c) and (d)). Whether ligation of both strands occurs on the same DNA molecule *in vivo* is uncertain. It is possible that on individual substrate molecules, only one of these strands is ligated. If so, a subsequent repair involving gap filling has to occur on the second strand by *E. coli* proteins. However, this scenario would lead to the potential problem that one Int-h monomer remains covalently linked *in cis* to the corresponding core

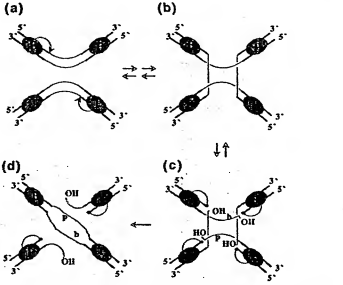


Figure 7. Schematic representation of one possible reaction pathway leading to deletion on pNC4. (a) *attL* and *attP* are depicted in antiparallel orientation within a hypothetical synaptic complex. The Int monomer (filled oval) bound to either the B or P arm initiates a nucleophilic attack on the top strand within each *att* site (indicated by curved arrows). (b) The first reciprocal pair of strand exchanges between top strands has been completed, leading to a Holliday junction intermediate structure. (c) All four Int-h (Int-h/218) monomers engage in a nucleophilic attack against their corresponding core sites and become covalently connected *in cis* through a 3′-phosphotyrosine linkage. This will lead to a double-stranded break within each *att* site. (c), (d) The 5′-OH from the overlap strand connected to the B arm in *attL* (labelled b) is rejoined with the P arm from *attP*. Likewise, the 5′-OH from the overlap strand which is still connected with the P arm in *attP* (labelled p) is ligated with the B arm in *attL*. This leads to the formation of a new hybrid *att* site with correct chemical polarity, but with reversed genetic polarity (i.e. BOP). Note that if both overlap strands, p and b, are religated within the same DNA molecule, the seven base-pair overlap region will adopt a heteroduplex structure containing five out of seven non-complementary bases. Based on the results presented in this study, the corresponding second hybrid *att* site (P′OP′) cannot be formed through religation. It is inferred that these DNAs which still carry Int monomers covalently linked at the 3′ ends will be degraded by *E. coli* proteins.

site and has to be removed prior to ligation. At the moment, we favor the possibility that both strands can be ligated on a single substrate molecule. This would lead to the formation of a seven base-pair heteroduplex structure in the overlap region of the new hybrid *att* site (Figure 7(d)), which could be resolved through repair and/or plasmid replication.

We have shown that the corresponding second hybrid *att* site cannot be generated by the mutant

Int. Based on our observation that the substrate DNA is lost with pNC5 in the presence of IHF, we think it is likely that *E. coli* enzymes degrade the linearized, origin-bearing DNA that contains Int-h monomers covalently linked to the 3′ ends (Figure 7(d)). This implies that double-stranded cleavage is efficient and eventually occurs on the entire population of pNC5 molecules inside *E. coli*. This, in turn, is in agreement with our observation of nearly 100% deletion on pNC4. Deletion on pNC2, on the other hand, is much less efficient. This can explain why pNC3 is stably propagated (Table 1). The fact that we were unable to identify the second hybrid *att* site suggests furthermore that strand ligation occurs within the framework of a synaptic complex, and not by random collision of freed DNA ends.

The role(s) of the accessory protein IHF in changing the directionality of strand exchange is puzzling. IHF is strictly required for deletion on pNC4 (Figure 5 and Table 1). One possibility is that additional Int-h (Int-h/218) monomers are delivered to the core sites of *attL* from its flanking arm binding sites *via* IHF-induced DNA-bending (Moitoso de Vargas *et al.*, 1988, 1989). Perhaps the presence of additional mutant Int monomers with an enhanced affinity for core sites interferes with the co-ordination of strand cleavage during isomerization of Holliday junctions. In contrast, deletion on pNC1 is only observed in the absence of IHF (Figure 3). IHF and Xis seem to direct Holliday junction resolution towards recombinant products in integrative and excisive recombination (Franz & Landy, 1995). It is possible that without these accessory proteins, the isomerization step required to switch to bottom strand exchange is impaired, so that the amount of this intermediate structure transiently increases. This, in turn, could result in a "cleavable synaptic complex" (Figure 7(c)) due to the presence of mutant instead of wild-type Int. A similar reasoning could account for deletion observed with substrates carrying inverted pairs of wild-type *attL* and *attR* (pNC2; Table 1). The situation is different again with pNC6 and pNC7, which require IHF for deletion (Figure 4 and Table 1). Since the overlap sequence in these pairs of *att* sites differs at position 0, it is likely that IHF is required to overcome an impairment in the isomerization step imposed by such a heterology. This again could lead to a transient increase in the amount of Holliday junctions, which are either resolved through reciprocal bottom strand exchange or cleaved and rejoined by mutant Int to yield deletion products.

The formation of so-called "contrary" recombinant products by wild-type Int has been observed before in *in vitro* studies using either heteroduplex or half-*att* site substrates (Nash & Robertson, 1989; Nunes-Düby *et al.*, 1989, 1997). These products, termed Y-structures, contain one recombinant strand that results from conventional top strand exchange, while the second strand shows normal chemical polarity but with reversed genetic polarity (e.g. BOP). Complete new hybrid *att* sites with reversed genetic polarity in both strands, as shown in the present study, were not observed. These Y-structures most likely result as a direct consequence of the aberrant structures of *att* sites. However, an important finding of these studies is that Int can join strands indiscriminately in the absence of complementary (homologous) strands.

The present study shows that true contrary recombinant products are generated *in vivo* with wild-type substrates for integrative and excisive recombination in the absence of IHF. It is important to note that these products are observed only with mutant Int. It is therefore an intrinsic property of Int-h and Int-h/218, and not that of a particular recombination substrate, that leads to the observed change in directionality of recombination. It is possible that the presence of a lysine residue instead of a glutamate residue at position 174 within the catalytic domain of Int somehow interferes with the normal communication between Int monomers within a Holliday junction. This could lead to an unco-ordinated nucleophilic attack on all four strands. Alternatively, or in addition, the presence of this lysine residue may interfere with the normal isomerization step of Holliday junctions, leading to a conformational change within this intermediate DNA structure that allows four Int monomers to attack the DNA backbones. An inspection of the recently solved structures of the Cre-*lox* recombination synapse and Holliday junction may be informative here (Gopaul *et al.*, 1998; Guo *et al.*, 1997). In the recombination synapse, alanine 131 and lysine 132 from the Cre subunit that has cleaved the *loxA* site are in contact with DNA. A sequence comparison of 105 members of the Int family shows that lysine 174 in Int-h aligns with alanine 131 in Cre (Nunes-Düby *et al.*, 1998). This is consistent with our hypothesis that Int-h (Int-h/218) interferes with the isomerization of Holliday junctions required to switch from top to bottom strand exchange, possibly through additional DNA contact(s) within this intermediate structure.

Materials and Methods

Bacterial strains

The following *E. coli* strains were employed in this study. DH5α (*supE44* Δ*lacU169* (Φ80*lacZ*ΔM15) *hsdR17* *recA1* *endA1* *gyrA96* *thi-1* *relA1*; Hanahan, 1983); DH1 (F⁻ *supE44* *recA1* *endA1* *gyrA96* *thi-1* *hsdR17* *relA1* λ⁻) (Hanahan, 1983); JC5547 (*thr-1* *ara-14* *leuB6* Δ(*gpt-proA*)*62* *lacY1* *tsx-33* *glnV44*(*AS*) *galK2*(Oc) λ⁻ *hisG4*(Oc) *recA13* *recC22* *recB21* *rpsL31*(*strR*) *xylA5* *mtl-1* *argE3*(Oc) *thi-1*; Willetts & Clark, 1969); CSH26 (F⁻ *ara*Δ(*lac pro*) *thi*; Miller, 1972); CSH26ΔIHF (F⁻ *ara*Δ(*lac pro*) *thi* *himA*Δ*82::Tn10*(*Tc*^{R}) *himD*Δ*3::cat*(*Cm*^{R}); kindly provided by B. Rak, Freiburg, Germany); CSH50 (F⁻ *ara*Δ(lac pro) thi strA; Miller, 1972); and CSH50ΔFis (F⁻ *ara*Δ(lac pro) thi strA fis::kan; Koch *et al.*, 1988).

Construction of integrase expression vectors

Int and Int-h were subcloned by PCR from pHN1 and pHN16 (Honigman *et al.*, 1979; Lange-Gustafson & Nash, 1984), respectively. Both genes were introduced into the polylinker from pTrc99A (Pharmacia) in which the *Nco*I site has been destroyed. Expression from pTrc99A is under the control of the strong trc promoter containing the *trp* (–35) and the *lac* UV5 (–10) region separated by 17 bp. Expression is regulated by the *lacI*^{q} gene product also encoded on the same plasmid. Both genes were amplified using the following primers: "IntproN1" which binds at the 5′ end (5′-GCTCTAGAATGGGAAGAAGGCGAAGTCA-3′) and "IntproCO" binding at the 3′ end (5′-AAAACTGCAGT CATTATTTGATTTCAATTTTGTCCC-3′). PCR products were resolved through agarose gel electrophoresis, isolated, and digested with *Xba*I and *Pst*I. The fragments were cloned into pTrc99A which was linearized by *Xba*I/*Pst*I, and the resulting expression vectors (pTrcInt and pTrcInt-h) amplified in DH5α.

The double mutant Int-h/218 was generated from pTrcInt-h by PCR-directed mutagenesis, which substituted the guanine at position 652 of the Int-h gene by an aderine residue. The resulting codon change at position 218 replaces the glutamate residue with a lysine residue. The following oligonucleotides were used as primers: nc1a218KR containing the altered nucleotide anneals at the coding strand of the Int-h gene (5′-GGTGATTTA TGCAAAATGAAGTGGTCTGATATCGTAGATGGA-3′) and nc1b218KL anneals at the non-coding strand and directs DNA synthesis in the opposite direction (5′-AACTCGTTGCCCGGTAACAACAGCCAGTTCCAT TGCAAG-3′). PCR was performed using the "Master Mix Kit" (Qiagen, Germany). The resulting linear expression vector was purified through agarose gel electrophoresis, isolated, and ligated after phosphorylation using T4 Ligase (New England Biolabs). Screening for a functional double mutant was performed through *in vivo* recombination assays (see below). The sequence of the sub-cloned Int, Int-h, and Int-h/218 gene was confirmed by DNA sequencing.

Construction of recombination substrates

Plasmids used as recombination substrates are listed in Table 1 and are derivatives of pACYC177 carrying the kanamycin resistance gene. pNC1 (9.13 kb) bearing wild-type *attP* and *attB* as inverted repeats was obtained by combining pAB3YC, a derivative of pAB3 (Dröge & Cozzarelli, 1989), with pACYC177. pAB3YC was partially digested with *Nhe*I in order to remove a 1.4 kb origin-containing fragment, and ligated with pACYC177 which was linearized at its unique *Nhe*I site. pNC2 (9.13 kb) was generated by *in vivo* recombination of pNC1 using wild-type Int in the presence of IHF (see below). pNC3 (9.13 kb) is a derivative of pNC2 in which the orientation of *attL* and *attR* with respect to both the origin and resistance gene has been changed. pNC4 (8.2 kb) was constructed by inserting a *Xmn*I-*Eco*RV *attP*-containing fragment from pNC1 into *Sca*I-linearized pACYC177. *attL* was obtained from pNC2 and subsequently inserted by blunt-end ligation into the *Eag*I site of the *attP*-bearing plasmid. pNC5 (8.59 kb) was derived from pNC2 by cloning *attP*, which is present on a *Nhe*I fragment, into a different position of *attR*-deleted pNC2. pNC6 (7.04 kb) was derived from pNC1 by replacing *attP* with *attP**. The modified *attP* site was generated by PCR-directed mutagenesis and first cloned into pTZ18R (Pharmacia). pNC7 (6.67 kb) is a derivative of pNC2 in which *attR* has been replaced by *attR**, the latter was obtained from *in vivo* recombination between a derivative of *attB*, termed *attH*, and *attP**. pNC8 (5.9 kb) is a derivative of pNC1, in which *attP* is replaced by *attR**. pNC9 (9.13 kb) is a derivative of pNC1, in which the orientation of *attP* has been inverted. pNC10 (9.13 kb) is a derivative of pNC2 in which *attR* has been inverted with respect to *attL*, so that both *att* sites are present as direct repeats. pNC11 (8.20 kb) was constructed as described for pNC4, but in this case selecting for the presence of *att* sites as direct repeats. pNC12 (7.04 kb) is a derivative of pNC1 in which *attP* was replaced by *attP** and screened for the desired orientation. pP′OP″ was generated by cloning two PCR-derived fragments into *Bam*HI/*Hind*III-cleaved pACYC184. For this, we used the unique DdeI restriction site present in *attP*. The sequence of the Int core binding sites and the overlap in P′OP″ is as follows: 5′-CAACTTAGTAT AAA-TAAGTTGGC-3′. It therefore represents one of two possible recombination products that were expected if deletion occurs on pNC5.

In vivo recombination

Expression vectors and recombination substrates were co-transformed into the appropriate *E. coli* strains mentioned in the text. After incubation of single colonies overnight in the presence of ampicillin to select for the expression vector and kanamycin to select for the substrate DNA, transformants were cultivated at 37°C for an additional 17 hours under selection pressure, and plasmid DNA isolated by affinity chromatography (Qiagen, Germany). Recombination was analyzed through restriction digests using the appropriate endonuclease (Table 1) and subsequent separation of DNA fragments through agarose gel electrophoresis. In order to test whether *recA*, *recB* or *recC* may be required for deletion on pNC4 and pNC6, we employed *E. coli* strains DH1 and JC5547. The requirement for Fis was tested by comparing deletion on a derivative of pNC6, which carries a spectomycin resistance gene, in CSH50 and CSH50ΔFis.

DNA sequencing

Deletion products were sequenced using the fluorescence-based 373A system (Applied Biosystems). The following two oligonucleotides were used as sequencing-primers: ATT-PC which anneals at the P arm in the direction of the overlap region (5′-TTGATAGCT CTTCCGCTTTCTGTTACAGGTCACTAATACC-3′), and ATT-BA which anneals at the complementary strand within the B arm (5′-GTCTAGCTAGCCGGGAAACTG AAAATGTGTTC-3′). The sequence of subcloned Int genes and that of Int-h/218 was determined using three oligonucleotides as primers. Two of them anneal within pTrc99A either upstream or downstream of the polylinker. The third anneals at nucleotide positions 331 to 348 within the Int gene.

Gel electrophoresis

DNA was analyzed through agarose gel electrophoresis (0.8%) in TBE buffer (90 mM Tris-borate (pH 8.3), 2.5 mM EDTA). DNA was visualized by UV after staining with ethidium bromide. Photographs were taken with the Image Master® System (Pharmacia).

Acknowledgments

Special thanks go to H. Nash, R. Kahmann, and B. Rak for plasmids and *E. coli* strains. Very special thanks are due to E. Lorbach for constructing pTrcInt-h and *attP**. We are grateful to B. Müller-Hill, A. Barker, and K. Schnetz for comments on the manuscript, and to A. Landy, G. Muskhelishvili, S. Nunes-Düby, and A. Segall for discussions. This work was financed through SFB 274 and DFG grant (Dr187/10-1). P.D. is a "Heisenberg" fellow of the DFG.

References

Azaro, A. M. & Landy, A. (1997). The isomeric preference of Holliday junctions influences resolution bias by λ integrase. *EMBO J.* 16, 3744-3755.

Better, M., Lu, C., Williams, R. C. & Echols, H. (1982). Site-specific DNA condensation and pairing mediated by the *Int* protein of bacteriophage λ. *Proc. Natl Acad. Sci. USA*, 79, 5837-5841.

Burgin, A. B. & Nash, H. A. (1992). Symmetry in the mechanism of bacteriophage λ integrative recombination. *Proc. Natl Acad. Sci. USA*, 89, 9642-9646.

Dorgai, L., Yagil, E. & Weisberg, R. A. (1995). Identifying determinants of recombination specificity: construction and characterization of mutant bacteriophage integrases. *J. Mol. Biol.* 252, 178-188.

Dröge, P. & Cozzarelli, N. R. (1989). Recombination of knotted substrates by Tn3 resolvase. *Proc. Natl Acad. Sci. USA*, 86, 6062-6066.

Franz, B. & Landy, A. (1990). Interactions between λ Int molecules bound to sites in the region of strand exchange are required for efficient Holliday junction resolution. *J. Mol. Biol.* 215, 523-535.

Franz, B. & Landy, A. (1995). The Holliday junction intermediates of λ integrative and excisive recombination respond differently to the bending proteins, IHF and Xis. *EMBO J.* 14, 397-406.

Goodman, S. D., Nicholson, S. C. & Nash, H. A. (1992). Deformation of DNA during site-specific recombination of bacteriophage λ: replacement of IHF protein by HU protein or sequence-directed bends. *Proc. Natl Acad. Sci. USA*, 89, 11910-11914.

Gopaul, D. N., Feng, G. & Van Duyne, G. D. (1998). Structure of the Holliday junction intermediate in Cre-*lox*P site-specific recombination. *EMBO J.* 17, 4175-4187.

Guo, K., Gopaul, D. N. & Van Duyne, G. D. (1997). Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature*, 389, 40-47.

Hallet, B. & Sherratt, D. J. (1997). Transposition and site-specific recombination: adapting DNA cut-and-paste mechanisms to a variety of genetic rearrangements. *FEMS Microbiol. Rev.* 21, 157-178.

Han, Y. W., Gumport, R. I. & Gardner, J. F. (1994). Mapping the functional domains of bacteriophage lambda integrase protein. *J. Mol. Biol.* 235, 908-925.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* 166, 557-580.

Honigman, A., Hu, S.-L. & Szybalski, W. (1979). Regulation of integration by coliphage λ: activation of *int* transcription by the cII and cIII proteins. *Virology*, 92, 542-556.

Kho, S. H. & Landy, A. (1994). Dissecting the resolution reaction of lambda integrase using suicide Holliday junction substrates. *EMBO J.* 13, 2714-2724.

Kim, S. & Landy, A. (1992). Lambda Int protein bridges between higher order complexes at two distant chromosomal loci *attL* and *attR*. *Science*, 256, 198-203.

Kitts, P. A. & Nash, H. A. (1987). Homology-dependent interactions in phage λ site-specific recombination. *Nature*, 329, 346-348.

Kitts, P. A. & Nash, H. A. (1988). Bacteriophage lambda site-specific recombination proceeds with a defined order of strand exchanges. *J. Mol. Biol.* 204, 95-107.

Koch, C., Vandekerckhove, J. & Kahmann, R. (1988). *Escherichia coli* host factor for site-specific DNA inversion: Cloning and characterization of the *fis* gene. *Proc. Natl Acad. Sci. USA*, 85, 4237-4241.

Kwon, H. J., Tirumalai, R., Landy, A. & Ellenberger, T. (1997). Flexibility in DNA recombination: structure of the λ integrase catalytic core. *Science*, 276, 126-131.

Landy, A. (1989). Dynamic, structural and regulatory aspects of lambda site-specific recombination. *Annu. Rev. Biochem.* 58, 913-949.

Landy, A. (1993). Mechanistic and structural complexity in the site-specific recombination pathways of Int and FLP. *Curr. Opin. Genet. Dev.* 3, 699-707.

Lange-Gustafson, B. J. & Nash, H. A. (1984). Purification and properties of Int-h, a variant protein involved in site-specific recombination by bacteriophage λ. *J. Biol. Chem.* 259, 12724-12732.

Miller, H. I., Mozola, L. A. & Friedman, D. I. (1980). *int*-h: an int mutation of phage λ that enhances site-specific recombination. *Cell*, 20, 721-729.

Miller, J. H. (1972). *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Moitoso de Vargas, L. & Landy, A. (1991). A switch in the formation of alternative DNA loops modulates λ site-specific recombination. *Proc. Natl Acad. Sci. USA*, 88, 588-592.

Moitoso de Vargas, L., Pargellis, C. A., Hasan, N. M., Bushman, E. W. & Landy, A. (1988). Autonomous DNA binding domains of λ integrase recognize different sequence families. *Cell*, 54, 923-929.

Moitoso de Vargas, L., Kim, S. & Landy, A. (1989). DNA looping generated by the DNA-bending protein IHF and the two domains of λ integrase. *Science*, 244, 1457-1461.

Nash, H. A. (1996). Site-specific recombination: integration, excision, resolution, and inversion of defined DNA segments. In Escherichia coli *and* Salmonella: *Cellular and Molecular Biology* (Neidhardt, F. C. *et al.*, eds), pp. 2363-2376, ASM Press, Washington, DC.

Nash, H. A. (1989). Heteroduplex substrates for bacteriophage lambda site-specific recombination: cleavage and strand transfer products. *EMBO J.* 8, 3523-3533.

Nunes-Düby, S. E., Matsumoto, L. & Landy, A. (1987). Site-specific recombination intermediates trapped with suicide substrates. *Cell*, 50, 779-788.

Nunes-Düby, S. E., Matsumoto, L. & Landy, A. (1989). Half-*att* site substrates reveal the homology independence and minimal protein requirements for productive synapsis in λ excisive recombination. *Cell*, 59, 197-206.

Nunes-Düby, S. E., Tirumalai, R. S., Dorgai, L., Yagil, E., Weisberg, R. A. & Landy, A. (1994). λ Integrase cleaves DNA *in cis*. *EMBO J.* 13, 4421-4430.

Nunes-Düby, S. E., Azaro, M. & Landy, A. (1995). Swapping DNA strands and sensing homology

without branch migration in λ site-specific recombination. *Curr. Biol.* **5**, 139-148.
Nunes-Düby, S. E., Yu, D. & Landy, A. (1997). Sensing homology and the strand swapping step in λ excisive recombination. *J. Mol. Biol.* **272**, 493-508.
Nunes-Düby, S. E., Tirumalai, R. S., Kwon, H. J., Ellenberger, T. & Landy, A. (1998). Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucl. Acids Res.* **26**, 391-406.
Patsey, R. L. & Bruist, M. L. (1995). Characterization of the interaction between the Lambda intasome and *attB*. *J. Mol. Biol.* **252**, 47-58.
Richet, E., Abcarian, P. & Nash, H. A. (1986). The interaction of recombination proteins with supercoiled DNA: defining the role of supercoiling in lambda integrative recombination. *Cell*, **46**, 1011-1021.
Richet, E., Abcarian, P. & Nash, H. A. (1988). Synapsis of attachment sites during lambda integrative recombination involves capture of a naked DNA by a protein-DNA complex. *Cell*, **52**, 9-17.
Sadowski, P. D. (1993). Site-specific genetic recombination: hips, flips, and flops. *FASEB J.* **7**, 760-767.
Segall, A. M. & Nash, H. A. (1993). Synaptic intermediates in bacteriophage lambda site-specific recombination: integrase can align pairs of attachment sites. *EMBO J.* **12**, 4567-4576.
Segall, A. M. & Nash, H. A. (1996). Architectural flexibility in lambda site-specific recombination: three alternate conformations channel the *attL* site into three distinct pathways. *Genes Cells*, **1**, 453-463.
Tirumalai, R. S., Healey, E. & Landy, A. (1997). The catalytic domain of λ site-specific recombinase. *Proc. Natl Acad. Sci. USA*, **94**, 6104-6109.
Tirumalai, R. S., Kwon, H. J., Healey, Cardente E., Ellenberger, T. & Landy, A. (1998). Recognition of core-type DNA sites by λ integrase. *J. Mol. Biol.* **279**, 513-527.
Willetts, N. S. & Clark, A. J. (1969). Characteristics of some multiply recombination-deficient strains of *Escherichia coli*. *J. Bacteriol.* **100**, 231-239.
Wu, Z., Gumport, R. I. & Gardner, J. F. (1997). Genetic analysis of second-site revertants of bacteriophage λ integrase mutants. *J. Bacteriol.* **179**, 4030-4038.
Yagil, E., Dorgai, L. & Weisberg, R. (1995). Identifying determinants of recombination specificity: construction and characterization of chimeric bacteriophage integrases. *J. Mol. Biol.* **252**, 163-177.
Yang, W. & Mizuuchi, K. (1997). Site-specific recombination in plane view. *Structure*, **5**, 1401-1406.

Edited by M. Yaniv

(Received 16 November 1998; received in revised form 22 March 1999; accepted 22 March 1999)

EXHIBIT 3

US005464764A

United States Patent [19]

Capecchi et al.

[11] Patent Number: 5,464,764

[45] Date of Patent: Nov. 7, 1995

[54] **POSITIVE-NEGATIVE SELECTION METHODS AND VECTORS**

[75] Inventors: **Mario R. Capecchi; Kirk R. Thomas,** both of Salt Lake City, Utah

[73] Assignee: **University of Utah Research Foundation,** Salt Lake City, Utah

[21] Appl. No.: **14,083**

[22] Filed: **Feb. 4, 1993**

Related U.S. Application Data

[63] Continuation of Ser. No. 397,707, Aug. 22, 1989, abandoned.

[51] **Int. Cl.**[6] **C12N 15/65**; C12N 15/64; C12N 15/09; C12N 15/11

[52] **U.S. Cl.** **435/172.3**; 435/172.1; 435/240.2; 435/320.1; 536/23.1; 536/23.2; 536/23.5; 536/23.6; 536/23.7; 536/23.72; 800/2; 800/DIG. 1; 800/DIG. 2; 935/22; 935/56; 935/70

[58] **Field of Search** ..... 800/2, 205, DIG. 1, 800/DIG. 2; 435/DIG. 67, 172.1–172.3, 320.1, 240.2, 240.1, 69.1, 6; 424/85.8, 85.91; 536/23.1, 23.2, 23.5, 23.6, 23.7, 23.72; 935/22, 56, 70

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,736,866	4/1988	Leder et al. .	
4,801,540	1/1989	Hiatt et al. .	
4,873,191	10/1989	Wagner et al. .	
4,916,213	4/1990	Scannon et al.	530/391
5,034,323	7/1991	Jorgensen et al.	435/172.3

FOREIGN PATENT DOCUMENTS

0247494	12/1987	European Pat. Off. .
0264166	4/1988	European Pat. Off. .
0279582	8/1988	European Pat. Off. .
WO8800239	1/1988	WIPO .
WO8801648	3/1988	WIPO .

OTHER PUBLICATIONS

"Transformation of Yeast", Hinnen, et al. (1978), *Proc. Natl. Acad. Sci. USA*, 75, 1929–1933.

"Insertion of DNA sequences into the human chromosomal β-globin locus by homologous recombination", Smithies, et al. (1985) *Nature*, 317, 230–234.

"Relative rates of homologous and nonhomologous recombination in transfected DNA", Roth, et al. (1985) *Proc. Natl. Acad. Sci. USA*, 82, 3355–3359.

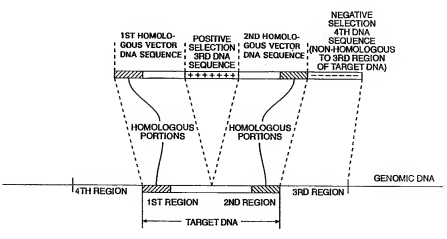
(List continued on next page.)

Primary Examiner—Jacqueline M. Stone
Assistant Examiner—Bruce Campbell
Attorney, Agent, or Firm—Townsend and Townsend Khourie and Crew

[57] **ABSTRACT**

Positive-negative selector (PNS) vectors are provided for modifying a target DNA sequence contained in the genome of a target cell capable of homologous recombination. The vector comprises a first DNA sequence which contains at least one sequence portion which is substantially homologous to a portion of a first region of a target DNA sequence. The vector also includes a second DNA sequence containing at least one sequence portion which is substantially homologous to another portion of a second region of a target DNA sequence. A third DNA sequence is positioned between the first and second DNA sequences and encodes a positive selection marker which when expressed is functional in the target cell in which the vector is used. A fourth DNA sequence encoding a negative selection marker, also functional in the target cell, is positioned 5' to the first or 3' to the second DNA sequence and is substantially incapable of homologous recombination with the target DNA sequence. The invention also includes transformed cells containing at least one predetermined modification of a target DNA sequence contained in the genome of the cell. In addition, the invention includes organisms such as non-human transgenic animals and plants which contain cells having predetermined modifications of a target DNA sequence in the genome of the organism.

44 Claims, 13 Drawing Sheets



OTHER PUBLICATIONS

"Mechanisms of Nonhomologous Recombination in Mammalian Cells", Roth, et al. (1985) *Mol. Cell. Biol.*, 5, 2599–2607.
"Gene targeting in plants", Paszkowski, et al. (1988), *EMBO J.*, 7, 4021–4026.
"Analysis of Homologous Recombination in Cultured Mammalian Cells", Folger, et al. (1984), *Cold Spring Harbor Symp. Quant. Biol.*, 49, 123–138.
"Homologous Recombination in Mouse L Cells", Lin, et al. (1984), *Cold Spring Harbor Symp. Quant. Biol.*, 49, 139–149.
"Homologous Recombination with DNA Introduced into Mammalian Cells", Smithies, et al. (1984), *Cold Spring Harbor Symp. Quant. Biol.*, 49, 161–170.
"Homologous Recombination between Defective neo Genes in Mouse 3T6 Cells", Smith, et al. (1984), *Cold Spring Harbor Symp. Quant. Biol.*, 49, 171–181.
"High Frequency Targeting of Genes to Specific Sites in the Mammalian Genome", Thomas, et al. (1986), *Cell*, 44, 419–428.
"Introduction of homologous DNA sequences into mammalian cells induces mutations in the cognate gene", Thomas, et al. (1986), *Nature*, 324, 34–38.
"Targeted correction of a mutant HPRT gene in mouse embryonic stem cells", Doetschman, et al. (1987), *Nature*, 330, 576–578.
"Accurate modification of a chromosomal plasmid by homologous recombination in human cells", Song, et al. (1987), *Proc. Natl. Acad. Sci. USA*, 84, 6820–6824.
"Site–Directed Mutagenesis by Gene Targeting in Mouse Embryo–Derived Stem Cells", Thomas, et al. (1987), *Cell*, 51, 503–512.
"Progress Toward Human Gene Therapy", Friedman, (1989), *Science*, 244, 1275–1281.
"Genetically Engineering Plants for Crop Improvement", Gasser, et al., (1989) *Science*, 244, 1293–1299.
"Genetic Engineering of Livestock", Pursel, et al., (1989) *Science*, 244, 1281–1288.
"Genetic Engineering of Filamentous Fungi", Timberlake, et al. (1989), *Science*, 244, 1313–1317.
"Positive genetic selection for gene disruption in mammalian cells by homologous recombination", Sedivy, et al. (1989), *Proc. Natl. Acad. Sci. USA*, 86, 227–231.
"Homologous integration in mammalian cells without target gene selection", Jasin, et al. (1988), *Genes and Development*, 2, 1353–1363.
"Recombinant fragment assay for gene targeting based on the polymerase chain reaction", Kim, et al. (1988), *Nucl. Acids Res.*, 16, 8887–8903.
"Production of chimaeric mice containing embryonic stem (ES) cells carrying a homoeobox Hox 1.1 allele mutated by homologous recombination", Zimmmer, et al. (1989), *Nature*, 338, 150–153.
"Production of a mutation in mouse En–2 gene by homologous recombination in embryonic stem cells", Joyner, et al. (1989), *Nature*, 33, 153–156.
"Transgenic Animals", Jaenisch (1988), *Science*, 240, 1468–1474.
"Implantation of Vascular Grafts Lines with Genetically Modified Endothelial Cells", Wilson, et al. (1989), *Science*, 244, 1344–1346.
"A large region (~95 kDa) of human factor VIII is dispensable for in vitro procoagulant activity", Toole, et al. (1986), *Proc. Natl. Acad. Sci. USA*, 83, 5939–5942 Paszkowski, et al. (1984), *EMBO J.*, 3, 2717–2722.
"Cloning of Large Segments of Exogenous DNA into Yeast by Means of Artificial Chromosome Vectors", Burke, et al. (1987), *Science*, 236, 806–812.
"Sequence homology requirements for intermolecular recombination in mammalian cells", Ayares, et al. (1986), *Proc. Natl. Acad. Sci. USA*, 83, 5199–5203.
"Altering the Genome by Homologous Recombination", Capecchi (1989), *Science*, 244, 1288–1292.
"Disruption of the proto–oncogene int–2 in mouse embryo–derived stem cells: a general strategy for targeting mutations to non–selectable genes", Mansour, et al. (1988), *Nature*, 336, 348–352.
"The pIC plasmid and phage vectors with versatile cloning sites for recombinant selection by insertional activation", Marsh, et al, (1984), *Gene*, 32, 481–485.
"The nucleotide sequence and transcript map of the herpes simplex virus thymidine kinase gene", McKnight (1980), *Nucl. Acids. Res.*, 8, 5949–5964.
"Nucleotide Sequence of the Herpes Simplex Virus Type 2 Thymidine Kinase Gene", Swain, et al. (1983), *J. Virol.*, 46, 1045–1050.
"A Foreign β–Globin Gene in Transgenic Mice: Integration at Abnormal Chromosomal Positions and Expression in Inappropriate Tissues", Lacy, et al. (1984) *Cell*, 34, 343–358.
"Plasmids Containing Mouse rDNA Do Not Recombine with Cellular Ribosomal Genes when Introduced into Cultured Mouse Cells", Steele, et al. (1984) *Molecular and Cell. Biol.*, 4, 576–582.
"Homologous and Nonhomologous Recombination in Monkey Cells", Subramani, et al. (1983) *Molecular and Cell. Biol.*, 3, 1040–1052.
"The Minimum Amount of Homology Required for Homologous Recombination in Mammalian Cells", Rubnitz, et al. (1984), *Molecular and Cell. Biol.*, 4, 2253–2258.
"Model for Homologous Recombination During Transfer of DNA into Mouse L Cells: Role for DNA Ends in the Recombination Process", Lin, et al. (1984) *Molecular and Cell. Biol.* 4, 1020–1034.
"Recombination in mouse L cells between DNA introduced into cells and homologous chromosomal sequences", Lin, et al. (1985), *Proc. Natl. Acad. Sci. USA*, 82, 1391–1395.
"HPRT–deficient (Lesch–Nyhan) mouse embryos derived from germline colonization by cultured cells", Hooper, et al. (1987), *Nature*, 326, 292–295.
"A potential animal model for Lesch–Nyhan syndrome through introduction of HPRT mutations into mice", Kuehn, et al. (1987), *Nature, 326, 295–298.*
"Homologous recombination between plasmids in mammalian cells can be enhanced by treatment of input DNA", Kucherlapati, et al. (1984) *Proc. Natl. Acad. Sci. USA*, 81, 3153–3157.
"Accurate modification of a chromosomal plasmid by homologous recombination in human cells", Song, et al. (1987), *Proc. Natl. Acad. Sci. USA*, 84, 6820–6824.
"Construction of Large DNA Segments in *Escherishia coli*", O'Connor, et al (1989), *Science*, 244, 1307–1312.
"Patterns of Integration of DNA Microinjected into Cultured Mammalian Cells: Evidence of Homologous Recombination Between Injected Plasmid DNA Molecules", Folger, K. R., et al. (1982), *Molecular and Cell. Biol.*, 2, 1372–1387.
"Nonreciprocal Exchanges of Information Between DNA Duplexes Coinjected into in Mammalian Cell Nuclei",

Folger, et al. (1985), *Molecular and Cell. Biol.*, 5, 59–69.
"Homologous Recombination between Coinjected DNA Sequences Peaks in Early to Mid–S Phase", Wong, et al. (1987), *Molecular and Cell. Biol.*, 7, 2294–2295.
"Targeting of Genes to Specific Sites in the Mammalian Genome", Thomas, et al. (1986), *Cold Spring Harbor Symposium on Quantitative Biology*, 51, 1101–1113.
"The new mouse genetics: Altering the Genome by Gene Targeting", Capecchi, (1989), *Trends in Genetics*, 5, 70–76.
"Introduction of homologous DNA sequences into mammalian cells induces mutations in the cognate gene", Thomas, et al. (1986), *Nature*, 324, 34–38.
"Altering the Genome by Homologous Recombination", Capecchi (1989), *Science*, 244, 1288–1292.
"High Frequency of Homologous Recombination in Mammalian Cells between Endogenous and Introduced SV40 Genomes", Jasin, et al. (1985), *Cell*, 43, 695–703.
"Rescue of Chromosomal T. Antigen Sequences into Extrachromosomally Replicating, Defective Simian Virus 40 DNA by Homologous Recombination", Subramani (1986), *Molecular and Cell. Biol.*, 6, 1320–1325.
Thompson, L., *The Washington Post, "New Method Raises Prospect for Altered–Gene Therapy," Nov. 24, 1988.*
Waldman, A. S., *Critical Reviews in Oncology Hematology*, 12: 49–64, 1992.
J M Sedivy et al (1992) Gene Targeting, pp. 160–162.
M Barinaga (1992) Science 257:1046–1047.
R M Strojek et al (1988) Genetic Engineering: Principles and Methods, pp. 239–240.

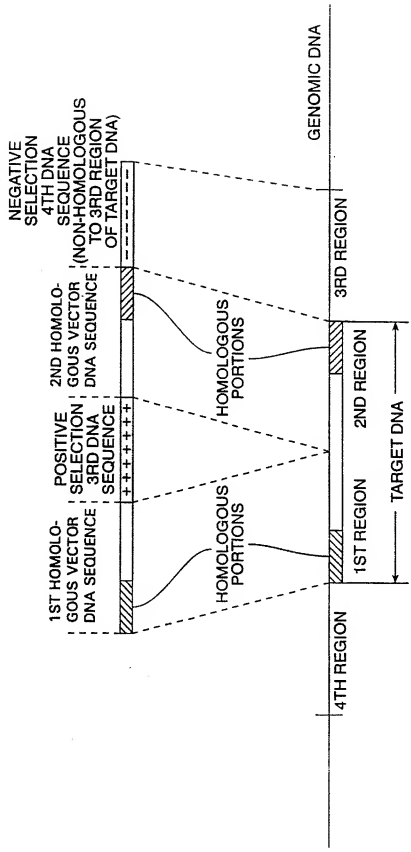

GENOMIC DNA
3RD REGION
2ND REGION
1ST REGION
4TH REGION
TARGET DNA
HOMOLOGOUS PORTIONS
HOMOLOGOUS PORTIONS
1ST HOMOLOGOUS VECTOR DNA SEQUENCE
POSITIVE SELECTION 3RD DNA SEQUENCE
2ND HOMOLOGOUS VECTOR DNA SEQUENCE
NEGATIVE SELECTION 4TH DNA SEQUENCE (NON-HOMOLOGOUS TO 3RD REGION OF TARGET DNA)

FIG. 1

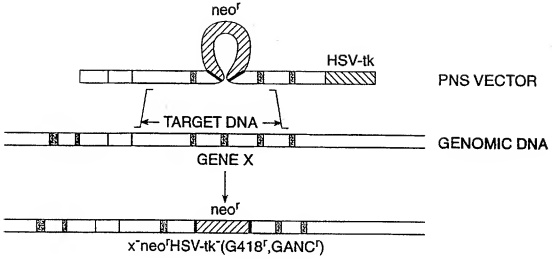

neo^r
HSV-tk
PNS VECTOR
TARGET DNA
GENOMIC DNA
GENE X
neo^r
x^-neo^rHSV-tk^-(G418^r,GANC^r)

FIG. 2A

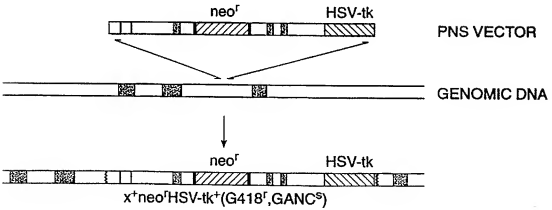

neo^r
HSV-tk
PNS VECTOR
GENOMIC DNA
neo^r
HSV-tk
x^+neo^rHSV-tk^+(G418^r,GANC^s)

FIG. 2B

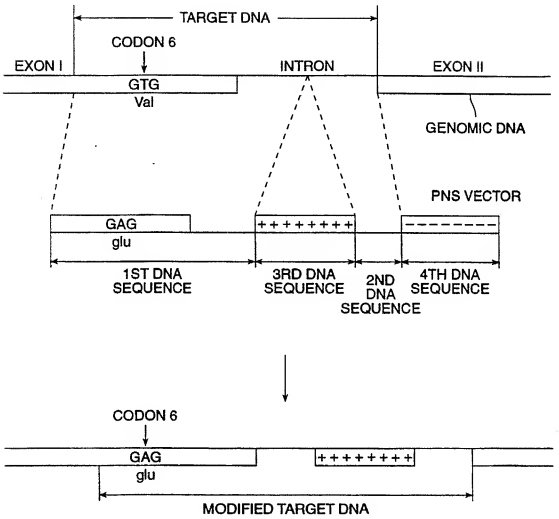

TARGET DNA
CODON 6
EXON I
INTRON
EXON II
GTG
Val
GENOMIC DNA
PNS VECTOR
GAG
glu
+++++++++
1ST DNA SEQUENCE
3RD DNA SEQUENCE
2ND DNA SEQUENCE
4TH DNA SEQUENCE
CODON 6
GAG
glu
+++++++++
MODIFIED TARGET DNA

FIG. 3

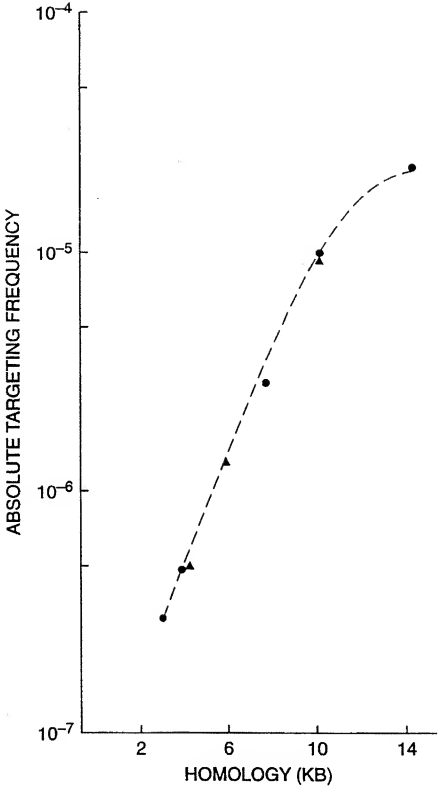

10^{-4}
10^{-5}
10^{-6}
10^{-7}
ABSOLUTE TARGETING FREQUENCY
2
6
10
14
HOMOLOGY (KB)

FIG. 4

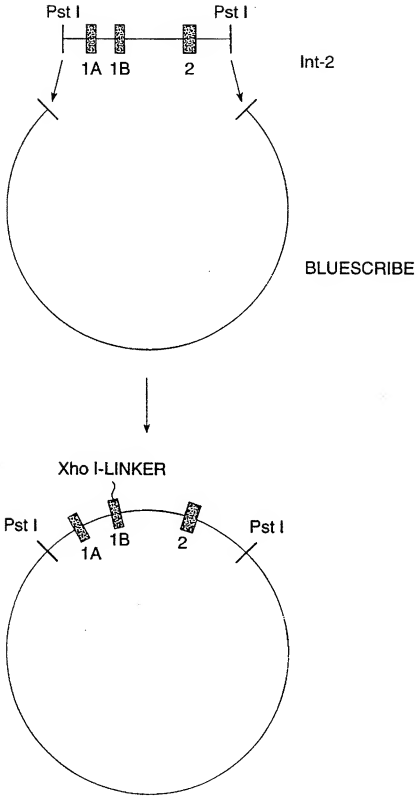
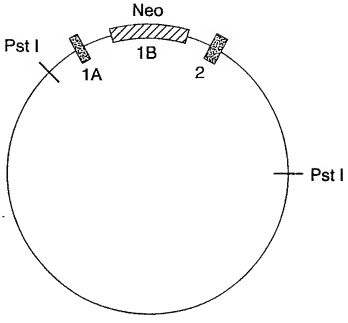

Pst I
Pst I
1A 1B
2
Int-2
BLUESCRIBE
Xho I-LINKER
Pst I
Pst I
1A
1B
2

FIG. 5A


Neo
Pst I
1A
1B
2
Pst I

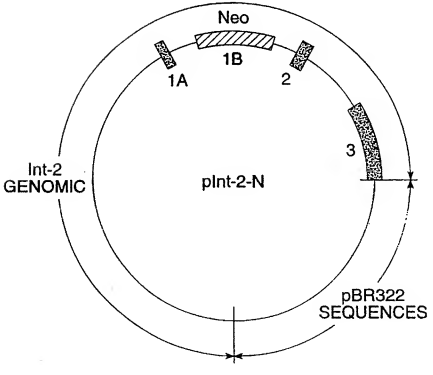

Neo
1A
1B
2
3
Int-2
GENOMIC
pInt-2-N
pBR322
SEQUENCES

FIG. 5B

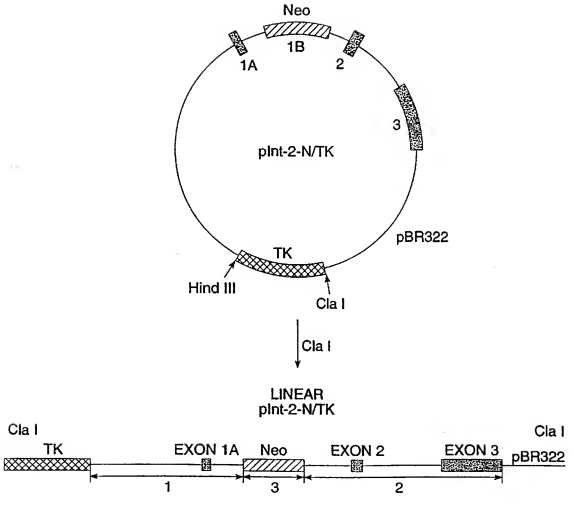
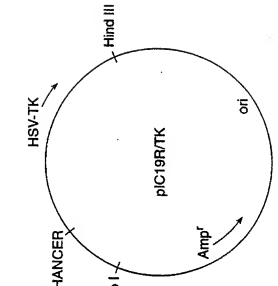

Neo
1B
1A
2
3
pInt-2-N/TK
pBR322
TK
Hind III
Cla I
Cla I
LINEAR
pInt-2-N/TK
Cla I
TK
EXON 1A
Neo
EXON 2
EXON 3
Cla I
pBR322
1
3
2

FIG. 5C



Met Thr Met Ile Thr Asn Ser Ile Ser Arg Ser Arg Ala Arg Glu Ser Leu Ala Ala Gly Arg Arg Ile Pro Gly Asn Ser Leu Ala
ATG ACC ATG ATT ACG AAT TCA ATA TCT AGA TCT CGA GCT CGC GAA AGC TTG GCT GCA GGT CGA CGG ATC CCC GGG AAT TCA CTG GCC

Eco RI Cla I Xba I Xho I Nru I Hind III Pst I Sal I BamH I Eco RI
Eco RV Bgl II Sac I Sma I

FIG. 5D

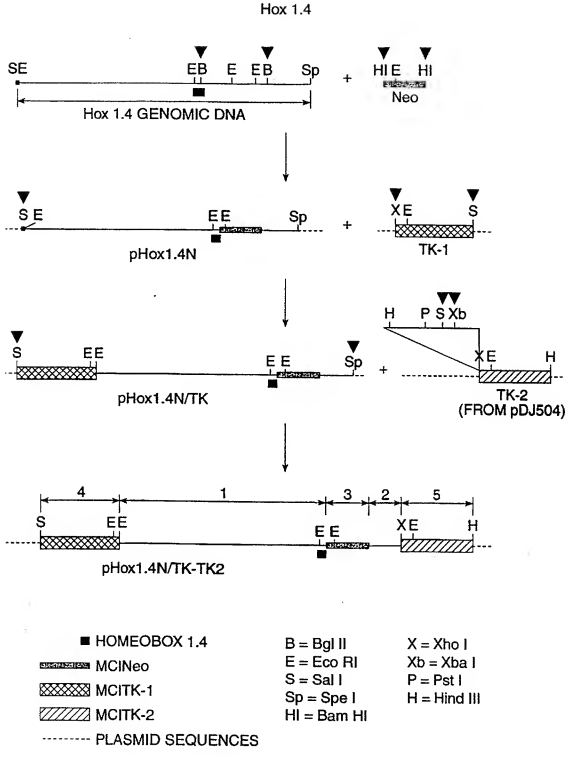

Hox 1.4
SE
EB E EB Sp
HI E HI
Neo
Hox 1.4 GENOMIC DNA
S E
E E
Sp
pHox1.4N
X E
S
TK-1
S
E E
E E
Sp
pHox1.4N/TK
H P S Xb
X E
H
TK-2
(FROM pDJ504)
4
1
3
2
5
S
E E
E E
X E
H
pHox1.4N/TK-TK2
HOMEOBOX 1.4
MCINeo
MCITK-1
MCITK-2
PLASMID SEQUENCES
B = Bgl II
E = Eco RI
S = Sal I
Sp = Spe I
HI = Bam HI
X = Xho I
Xb = Xba I
P = Pst I
H = Hind III

FIG. 6

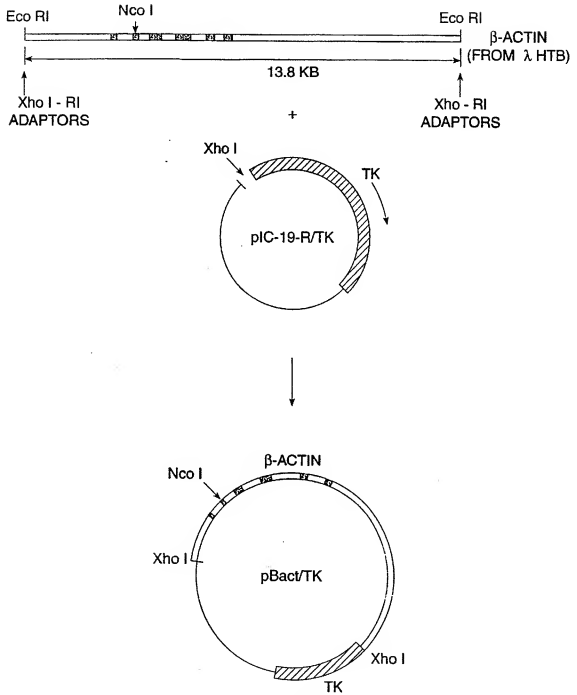

Eco RI
Nco I
Eco RI
β-ACTIN
(FROM λ HTB)
13.8 KB
Xho I - RI
ADAPTORS
+
Xho - RI
ADAPTORS
Xho I
TK
pIC-19-R/TK
β-ACTIN
Nco I
Xho I
pBact/TK
Xho I
TK

FIG. 7A

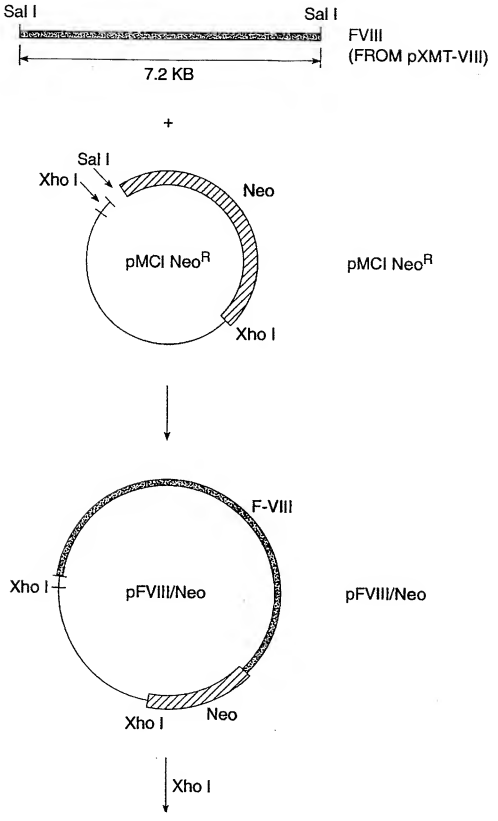

Sal I
Sal I
FVIII
(FROM pXMT-VIII)
7.2 KB
+
Sal I
Xho I
Neo
pMCI Neo^R
Xho I
pMCI Neo^R
F-VIII
Xho I
pFVIII/Neo
pFVIII/Neo
Xho I
Neo
Xho I

FIG. 7B

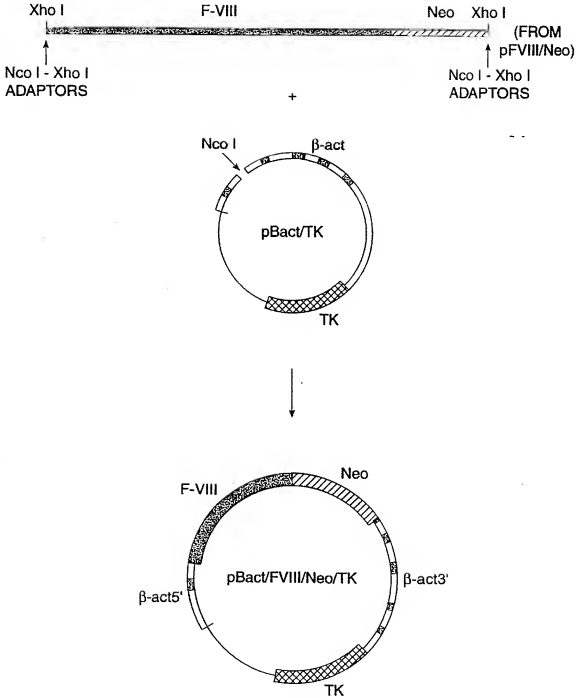

Xho I
F-VIII
Neo
Xho I
(FROM pFVIII/Neo)
Nco I - Xho I ADAPTORS
Nco I - Xho I ADAPTORS
+
Nco I
β-act
pBact/TK
TK
F-VIII
Neo
pBact/FVIII/Neo/TK
β-act5'
β-act3'
TK

FIG. 7C

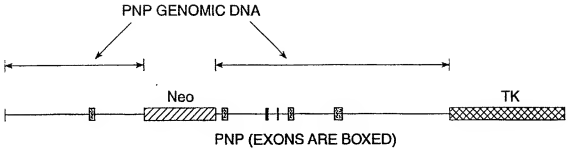


FIG. 8

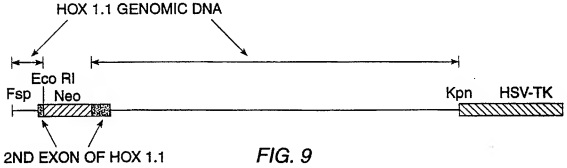


FIG. 9

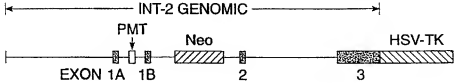


FIG. 10

POSITIVE-NEGATIVE SELECTION METHODS AND VECTORS

This invention was funded under grant No. R01-GM-21168 issued by the U.S. Department of Health and Human Services. This is a continuation of application Ser. No. 07/397,707, filed Aug. 22, 1989, now abandoned.

TECHNICAL FIELD OF THE INVENTION

The invention relates to cells and non-human organisms containing predetermined genomic modifications of the genetic material contained in such cells and organisms. The invention also relates to methods and vectors for making such modifications.

BACKGROUND OF THE INVENTION

Many unicellular and multicellular organisms have been made containing genetic material which is not otherwise normally found in the cell or organism. For example, bacteria, such as *E. coli*, have been transformed with plasmids which encode heterologous polypeptides, i.e., polypeptides not normally associated with that bacterium. Such transformed cells are routinely used to express the heterologous gene to obtain the heterologous polypeptide. Yeasts, filamentous fungi and animal cells have also been transformed with genes encoding heterologous polypeptides. In the case of bacteria, heterologous genes are readily maintained by way of an extra chromosomal element such as a plasmid. More complex cells and organisms such as filamentous fungi, yeast and mammalian cells typically maintain the heterologous DNA by way of integration of the foreign DNA into the genome of the cell or organism. In the case of mammalian cells and most multicellular organisms such integration is most frequently random within the genome.

Transgenic animals containing heterologous genes have also been made. For example, U.S. Pat. No. 4,736,866 discloses transgenic non-human mammals containing activated oncogenes. Other reports for producing transgenic animals include PCT Publication No. W082/04443 (rabbit β-globin gene DNA fragment injected into the pronucleus of a mouse zygote); EPO Publication No. 0 264 166 (Hepatitis B surface antigen and tPA genes under control of the whey acid protein promotor for mammary tissue specific expression); EPO Publication No. 0 247 494 (transgenic mice containing heterologous genes encoding various forms of insulin); PCT Publication No. W088/00239 (tissue specific expression of a transgene encoding factor IX under control of a whey protein promotor); PCT Publication No. W088/01648 (transgenic mammal having mammary secretory cells incorporating a recombinant expression system comprising a mammary lactogen-inducible regulatory region and a structural region encoding a heterologous protein); and EPO Publication No. 0 279 582 (tissue specific expression of chloramphenicol acetyltrans-ferase under control of rat β-casein promotor in transgenic mice). The methods and DNA constructs ("transgenes") used in making these transgenic animals also result in the random integration of all or part of the transgene into the genome of the organism. Typically, such integration occurs in an early embryonic stage of development which results in a mosaic transgenic animal. Subsequent generations can be obtained, however, wherein the randomly inserted transgene is contained in all of the somatic cells of the transgenic animals.

Transgenic plants have also been produced. For example, U.S. Pat. No. 4,801,540 to Hiatt, et al., discloses the transformation of plant cells with a plant expression vector containing tomato polygalacturonase (PG) oriented in the opposite orientation for expression. The anti-sense RNA expressed from this gene is capable of hybridizing with endogenous PG mRNA to suppress translation. This inhibits production of PG and as a consequence the hydrolysis of pectin by PG in the tomato.

While the integration of heterologous DNA into cells and organisms is potentially useful to produce transformed cells and organisms which are capable of expressing desired genes and/or polypeptides, many problems are associated with such systems. A major problem resides in the random pattern of integration of the heterologous gene into the genome of cells derived from multicellular organisms such as mammalian cells. This often results in a wide variation in the level of expression of such heterologous genes among different transformed cells. Further, random integration of heterologous DNA into the genome may disrupt endogenous genes which are necessary for the maturation, differentiation and/or viability of the cells or organism. In the case of transgenic animals, gross abnormalities are often caused by random integration of the transgene and gross rearrangements of the transgene and/or endogenous DNA often occur at the insertion site. For example, a common problem associated with transgenes designed for tissue-specific expression involves the "leakage" of expression of the transgenes. Thus, transgenes designed for the expression and secretion of a heterologous polypeptide in mammary secretory cells may also be expressed in brain tissue thereby producing adverse effects in the transgenic animal. While the reasons for transgene "leakage" and gross rearrangements of heterologous and endogenous DNA are not known with certainty, random integration is a potential cause of expression leakage.

One approach to overcome problems associated with random integration involves the use gene of targeting. This method involves the selection for homologous recombination events between DNA sequences residing in the genome of a cell or organism and newly introduced DNA sequences. This provides means for systematically altering the genome of the cell or organism.

For example, Hinnen, J. B., et al. (1978) *Proc. Natl. Acad. Sci. U.S.A.*, 75, 1929–1933 report homologous recombination between a leu2⁻ plasmid and a leu2⁻ gene in the yeast genome. Successful homologous transformants were positively selected by growth on media deficient in leucine.

For mammalian systems, several laboratories have reported the insertion of exogenous DNA sequences into specific sites within the mammalian genome by way of homologous recombination. For example, Smithies, O., et al. (1985) *Nature*, 317, 230–234 report the insertion of a linearized plasmid into the genome of cultured mammalian cells near the β-globin gene by homologous recombination. The modified locus so obtained contained inserted vector sequences containing a neomycin resistance gene and a sup F gene encoding an amber suppressor t-RNA positioned between the δ and β-globin structural genes. The homologous insertion of this vector also resulted in the duplication of some of the DNA sequence between the δ and β-globin genes and part of the β-globin gene itself. Successful transformants were selected using a neomycin related antibiotic. Since most transformation events randomly inserted this plasmid, insertion of this plasmid by homologous recombination did not confer a selectable, cellular phenotype for homologous recombination mediated transformation. A laborious screening test for identifying predicted targeting events using plasmid rescue of the supF marker in a phage

library prepared from pools of transfected colonies was used. Sib selection utilizing this assay identified the transformed cells in which homologous recombination had occurred.

A significant problem encountered in detecting and isolating cells, such as mammalian and plant cells, wherein homologous recombination events have occurred lies in the greater propensity for such cells to mediate non-homologous recombination. See Roth, D. B., et al. (1985) *Proc. Natl. Acad. Sci. U.S.A.*, 82 3355–3359; Roth, D. B., et al. (1985), *Mol. Cell. Biol.*, 5, 2599–2607; and Paszkowski, J., et al. (1988), *EMBO J.*, 7, 4021–4026. In order to identify homologous recombination events among the vast pool of random insertions generated by non-homologous recombination, early gene targeting experiments in mammalian cells were designed using cell lines carrying a mutated form of either a neomycin resistance (neo^r) gene or a herpes simplex virus thymidine kinase (HSV-tk) gene, integrated randomly into the host genome. Such exogenous defective genes were then specifically repaired by homologous recombination with newly introduced exogenous DNA carrying the same gene bearing a different mutation. Productive gene targeting events were identified by selection for cells with the wild type phenotype, either by resistance to the drug G418 (neo^r) or ability to grow in HAT medium (tk^+). See, e.g., Folger, K. R., et al. (1984), *Cold Spring Harbor Symp. Quant. Biol.*, 49, 123–138; Lin, F. L. et al. (1984), *Cold Spring Harbor Symp. Quant. Biol.*, 49, 139–149; Smithies, O., et al. (1984), *Cold Spring Harbor Symp. Quant. Biol.*, 49, 161–170; Smith, A. J. H., et al. (1984), *Cold Spring Harbor Symp. Quant. Biol.*, 49, 171–181; Thomas K. R., et al. (1986), *Cell, 41, 419–428;* Thomas, K. R., et al. (1986), *Nature*, 324, 34–38; Doetschman, T., et al. (1987), Nature, 330, 576–578; and Song, Kuy-Young, et al. (1987), *Proc. Natl. Acad. Sci. U.S.A.*, 84, 6820–6824. A similar approach has been used in plant cells where partially deleted neomycin resistance genes reportedly were randomly inserted into the genome of tobacco plants. Transformation with vectors containing the deleted sequences conferred resistance to neomycin in those plant cells wherein homologous recombination occurred. Paszkowski, J., et al. (1988), *EMBO J.*, 7, 4021–4026.

A specific requirement and significant limitation to this approach is the necessity that the targeted gene confer a positive selection characteristic in those cells wherein homologous recombination has occurred. In each of the above cases, a defective exogenous positive selection marker was inserted into the genome. Such a requirement severely limits the utility of such systems to the detection of homologous recombination events involving inserted selectable genes.

In a related approach, Thomas, K. R., et al. (1987), *Cell*, 51, 503–512, report the disruption of a selectable endogenous mouse gene by homologous recombination. In this approach, a vector was constructed containing a neomycin resistance gene inserted into sequences encoding an exon of the mouse hypoxanthine phosphoribosyl transferase (Hprt) gene. This endogenous gene was selected for two reasons. First, the Hprt gene lies on the X-chromosome. Since embryonic stem cells (ES cells) derived from male embryos are hemizygous for Hprt, only a single copy of the Hprt gene need be inactivated by homologous recombination to produce a selectable phenotype. Second, selection procedures are available for isolating $Hprt^-$ mutants. Cells wherein homologous recombination events occurred could thereafter be positively selected by detecting cells resistant to neomycin (neo^R) and 6-thioguanine ($Hprt^-$).

A major limitation in the above methods has been the requirement that the target sequence in the genome, either endogenous or exogenous, confer a selection characteristic to the cells in which homologous recombination has occurred (i.e. neo^R, tk^+ or $Hprt^-$). Further, for those gene sequences which confer a selectable phenotype upon homologous recombination (e.g. the Hprt gene), the formation of such a selectable phenotype requires the disruption of the endogenous gene.

The foregoing approaches to gene targeting are clearly not applicable to many emerging technologies. See, e.g. Friedman, T. (1989), *Science*, 244, 1275–1281 (human gene therapy); Gasser, C. S., et al., *Id.*, 1293–1299 (genetic engineering of plants); Pursel, I. G., et al., *Id.*. 1281–1288 (genetic engineering of livestock); and Timberlake, W. E., et al., *Id.* et al., 13—13, 1312 (genetic engineering of filamentous fungi). Such techniques are generally not useful to isolate transformants wherein non-selectable endogenous genes are disrupted or modified by homologous recombination. The above methods are also of little or no use for gene therapy because of the difficulty in selecting cells wherein the genetic defect has been corrected by way of homologous recombination.

Recently, several laboratories have reported the expression of an expression-defective exogenous selection marker after homologous integration into the genome of mammalian cells. Sedivy, J. M., et al. (1989), *Proc. Nat. Acad. Sci. U.S.A.*, 86, 227–231, report targeted disruption of the hemizygous polyomavirus middle-T antigen with a neomycin resistance gene lacking an initiation codon. Successful transformants were selected for resistance to G418. Jasin, M., et al. (1988), *Genes and Development*, 2, 1353–1363 report integration of an expression-defective gpt gene lacking the enhancer in its SV40 early promotor into the SV40 early region of a gene already integrated into the mammalian genome. Upon homologous recombination, the defective gpt gene acts as a selectable marker.

Assays for detecting homologous recombination have also recently been reported by several laboratories. Kim, H. S., et al. (1988), *Nucl. Acid. S. Res.*, 16, 8887–8903, report the use of the polymerase chain reaction (PCR) to identify the disruption of the mouse hprt gene. A similar strategy has been used by others to identify the disruption of the Hox 1.1 gene in mouse ES cells (Zimmmer, A. P., et al. (1989), *Nature*, 338, 150–153) and the disruption of the En-2 gene by homologous recombination in embryonic stem cells. (Joyner, A. L., et al. (1989), *Nature*, 338, 153–156).

It is an object herein to provide methods whereby any predetermined region of the genome of a cell or organism may be modified and wherein such modified cells can be selected and enriched.

It is a further object of the invention to provide novel vectors used in practicing the above methods of the invention.

Still further, an object of the invention is to provide transformed cells which have been modified by the methods and vectors of the invention to contain desired mutations in specific regions of the genome of the cell.

Further, it is an object herein to provide non-human transgenic organisms, which contain cells having predetermined genomic modifications.

The references discussed above are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the inventors are not entitled to antedate such disclosure by virtue of prior invention.

SUMMARY OF THE INVENTION

In accordance with the above objects, positive-negative selector (PNS) vectors are provided for modifying a target DNA sequence contained in the genome of a target cell capable of homologous recombination. The vector comprises a first DNA sequence which contains at least one sequence portion which is substantially homologous to a portion of a first region of a target DNA sequence. The vector also includes a second DNA sequence containing at least one sequence portion which is substantially homologous to another portion of a second region of a target DNA sequence. A third DNA sequence is positioned between the first and second DNA sequences and encodes a positive .selection marker which when expressed is functional in the target cell in which the vector is used. A fourth DNA sequence encoding a negative selection marker, also functional in the target cell, is positioned 5' to the first or 3' to the second DNA sequence and is substantially incapable of homologous recombination with the target DNA sequence.

The above PNS vector containing two homologous portions and a positive and a negative selection marker can be used in the methods of the invention to modify target DNA sequences. In this method, cells are first transfected with the above vector. During this transformation, the PNS vector is most frequently randomly integrated into the genome of the cell. In this case, substantially all of the PNS vector containing the first, second, third and fourth DNA sequences is inserted into the genome. However, some of the PNS vector is integrated into the genome via homologous recombination. When homologous recombination occurs between the homologous portions of the first and second DNA sequences of the PNS vector and the corresponding homologous portions of the endogenous target DNA of the cell, the fourth DNA sequence containing the negative selection marker is not incorporated into the genome. This is because the negative selection marker lies outside of the regions of homology in the endogenous target DNA sequence. As a consequence, at least two cell populations are formed. That cell population wherein random integration of the vector has occurred can be selected against by way of the negative selection marker contained in the fourth DNA sequence. This is because random events occur by integration at the ends of linear DNA. The other cell population wherein gene targeting has occurred by homologous recombination are positively selected by way of the positive selection marker contained in the third DNA sequence of the vector. This cell population does not contain the negative selection marker and thus survives the negative selection. The net effect of this positive-negative selection method is to substantially enrich for transformed cells containing a modified target DNA sequence.

If in the above PNS vector, the third DNA sequence containing the positive selection marker is positioned between first and second DNA sequences corresponding to DNA sequences encoding a portion of a polypeptide (e.g. within the exon of a eucaryotic organism) or within a regulatory region necessary for gene expression, homologous recombination allows for the selection of cells wherein the gene containing such target DNA sequences is modified such that it is non functional.

If, however, the positive selection marker contained in the third DNA sequence .of the PNS vector is positioned within an untranslated region of the genome, e.g. within an intron in a eucaryotic gene, modifications of the surrounding target sequence (e.g. exons and/or regulatory regions) by way of substitution, insertion and/or deletion of one or more nucleotides may be made without eliminating the functional character of the target gene.

The invention also includes transformed cells containing at least one predetermined modification of a target DNA sequence contained in the genome of the cell.

In addition, the invention includes organisms such as non-human transgenic animals and plants which contain cells having predetermined modifications of a target DNA sequence in the genome of the organism.

Various other aspects of the invention will be apparent from the following detailed description, appended drawings and claims.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 depicts the positive-negative selection (PNS) vector of the invention and a target DNA sequence.

FIGS. 2A and **2B** depict the results of gene targeting (homologous recombination) and random integration of a PNS vector into a genome respectively.

FIG. 3 depicts a PNS vector containing a positive selection marker within a sequence corresponding, in part, to an intron of a target DNA sequence.

FIG. 4 is a graphic representation of the absolute frequency of homologous recombination versus the amount of 100% sequence homology in the first and second DNA sequences of the PNS vectors of the invention.

FIGS. 5A, 5B, 5C and **5D** depict the construction of a PNS vector used to disrupt the INT-2 gene.

FIG. 6 depicts the construction of a PNS vector used to disrupt the HOX1.4 gene.

FIGS. 7A, 7B and **7C** depict the construction of a PNS vector used to transform endothelial cells to express factor VIII.

FIG. 8 depicts a PNS vector to correct a defect in the purine nucleoside phosphorylase gene.

FIG. 9 depicts a vector for promoterless PNS.

FIG. 10 depicts the construction of a PNS vector to target an inducible promoter into the int-2 locus.

DETAILED DESCRIPTION OF THE INVENTION

The positive-negative selection ("PNS") methods and vectors of the invention are used to modify target DNA sequences in the genome of cells capable of homologous recombination.

A schematic diagram of a PNS vector of the invention is shown in **FIG. 1**. As can be seen, the PNS vector comprises four DNA sequences. The first and second DNA sequences each contain portions which are substantially homologous to corresponding homologous portions in first and second regions of the targeted DNA. Substantial homology is necessary between these portions in the PNS vector and the target DNA to insure targeting of the PNS vector to the appropriate region of the genome.

As used herein, a "target DNA sequence" is a predetermined region within the genome of a cell which is targeted for modification by the PNS vectors of the invention. Target DNA sequences include structural genes (i.e., DNA sequences encoding polypeptides including in the case of eucaryots, introns and exons), regulatory sequences such as enhancers sequences, promoters and the like and other regions within the genome of interest. A target DNA sequence may also be a sequence which, when targeted by

a vector has no effect on the function of the host genome. Generally, the target DNA contains at least first and second regions. See FIG. 1. Each region contains a homologous sequence portion which is used to design the PNS vector of the invention. In some instances, the target DNA sequence also includes a third and in some cases a third and fourth region. The third and fourth regions are substantially contiguous with the homologous portions of the first and second region. The homologous portions of the target DNA are homologous to sequence portions contained in the PNS vector. The third and in some cases third and fourth regions define genomic DNA sequences within the target DNA sequence which are not substantially homologous to the fourth and in some cases fourth and fifth DNA sequences of the PNS vector.

Also included in the PNS vector are third and fourth DNA sequences which encode respectively "positive" and "negative" selection markers. Examples of preferred positive and negative selection markers are listed in Table I. The third DNA sequence encoding the positive selection marker is positioned between the first and second DNA sequences while the fourth DNA sequence encoding the negative selection marker is positioned either 3' to the second DNA sequences shown in FIG. 1, or 5' to the first DNA sequence (not shown in FIG. 1). The positive and negative selection markers are chosen such that they are functional in the cells containing the target DNA.

Positive and/or negative selection markers are "functional" in transformed cells if the phenotype expressed by the DNA sequences encoding such selection markers is capable of conferring either a positive or negative selection characteristic for the cell expressing that DNA sequence. Thus, "positive selection" comprises contacting cells transfected with a PNS vector with an appropriate agent which kills or otherwise selects against cells not containing an integrated positive selection marker. "Negative selection" on the other hand comprises contacting cells transfected with the PNS vector with an appropriate agent which kills or otherwise selects against cells containing the negative selection marker. Appropriate agents for use with specific positive and negative selection markers and appropriate concentrations are listed in Table I. Other positive selection markers include DNA sequences encoding membrane bound polypeptides. Such polypeptides are well known to those skilled in the art and contain a secretory sequence, an extracellular domain, a transmembrane domain and an intracellular domain. When expressed as a positive selection marker, such polypeptides associate with the target cell membrane. Fluorescently labelled antibodies specific for the extracellular domain may then be used in a fluoresence activated cell sorter (FACS) to select for cells expressing the membrane bound polypeptide. FACS selection may occur before or after negative selection.

TABLE I

Selectable Markers for Use in PNS-Vectors

Gene	Type	Selective Agents	Preferred Concentration of selective Agent	Organism
Neo	+	G418	50–1000 μg/ml	Eukaryotes
Neo	+	Kanamycin	5–500 μg/ml	Plants
Hyg	+	Hygromycin	10–1000	Eukaryotes

TABLE I-continued

Selectable Markers for Use in PNS-Vectors

Gene	Type	Selective Agents	Preferred Concentration of selective Agent	Organism
hisD	+	Histidinol	μg/ml 5–500 μg/ml	Animals
Gpt	+	Xanthine,	50–500 μg/ml	Animals
Ble	+	Bleomycin	1–100 μg/ml	Plants
Hprt	+	Hypoxanthine	0.01–10 mM	All
HSV-tk	–	Acyclovir	1–100 μM	Animals
		Gancyclovir	0.05–200 μM	Animals
		FIAU	0.02–100 μM	Animals
Hprt	–	6-thioguanine	0.1–100 μg/ml	All
Gpt	–	6-thioxanthine	0.1–100 μg/ml	Animals
Diphtheria toxin	–	None	None	Animals
Ricin toxin	–	None	None	Animals
cytosine deaminase	–	5-fluoro-cytosine	10–500 μg/ml	All

The expression of the negative selection marker in the fourth DNA sequence is generally under control of appropriate regulatory sequences which render its expression in the target cell independent of the expression of other sequences in the PNS vector or the target DNA. The positive selection marker in the third DNA, however, may be constructed so that it is independently expressed (eg. when contained in an intron of the target DNA) or constructed so that homologous recombination will place it under control of regulatory sequences in the target DNA sequence. The strategy and details of the expression of the positive selection marker will be discussed in more detail hereinafter.

The positioning of the negative selection marker as being either "5" or "3" is to be understood as relating to the positioning of the negative selection marker relative to the 5' or 3' end of one of the strands of the double-stranded PNS vector. This should be apparent from FIG. 1. The positioning of the various DNA sequences within the PNS vector, however, does not require that each of the four DNA sequences be transcriptionally and translationally aligned on a single strand of the PNS vector. Thus, for example, the first and second DNA sequences may have a 5' to 3' orientation consistent with the 5' to 3' orientation of regions 1 and 2 in the target DNA sequence. When so aligned, the PNS vector is a "replacement PNS vector" upon homologous recombination the replacement PNS vector replaces the genomic DNA sequence between the homologous portions of the target DNA with the DNA sequences between the homologous portion of the first and second DNA sequences of the PNS vector. Sequence replacement vectors are preferred in practicing the invention. Alternatively, the homologous portions of the first and second DNA sequence in the PNS vector may be inverted relative to each other such that the homologous portion of DNA sequence 1 corresponds 5' to 3' with the homologous portion of region 1 of the target DNA sequence whereas the homologous portion of DNA sequence 2 in the PNS vector has an orientation which is 3' to 5' for the homologous portion of the second region of the second region of the target DNA sequence. This inverted orientation

provides for and "insertion PNS vector". When an insertion PNS vector is homologously inserted into the target DNA sequence, the entire PNS vector is inserted into the target DNA sequence without replacing the homologous portions in the target DNA. The modified target DNA so obtained necessarily contains the duplication of at least those homologous portions of the target DNA which are contained in the PNS vector. Sequence replacement vectors and sequence insertion vectors utilizing a positive selection marker only are described by Thomas et al. (1987), *Cell*, 51, 503–512.

Similarly, the third and fourth DNA sequences may be transcriptionally inverted relative to each other and to the transcriptional orientation of the target DNA sequence. This is only the case, however, when expression of the positive and/or negative selection marker in the third and/or fourth DNA sequence respectively is independently controlled by appropriate regulatory sequences. When, for example a promoterless positive selection marker is used as a third DNA sequence such that its expression is to be placed under control of an endogenous regulatory region, such a vector requires that the positive selection marker be positioned so that it is in proper alignment (5' to 3' and proper reading frame) with the transcriptional orientation and sequence of the endogenous regulatory region.

Positive-negative selection requires that the fourth DNA sequence encoding the negative marker be substantially incapable of homologous recombination with the target DNA sequence. In particular, the fourth DNA sequence should be substantially non-homologous to a third region of the target DNA. When the fourth DNA sequence is positioned 3' to the second DNA sequence, the fourth DNA sequence is non-homologous to a third region of the target DNA which is adjacent to the second region of the target DNA. See FIG. 1. When the fourth DNA sequence is located 5' to the first DNA sequence, it is non-homologous to a fourth region of the target DNA sequence adjacent to the first region of the target DNA.

In some cases, the PNS vector of the invention may be constructed with a fifth DNA sequence also encoding a negative selection marker. In such cases, the fifth DNA sequence is positioned at the opposite end of the PNS vector to that containing the fourth DNA sequence. The fourth DNA sequence is substantially non-homologous to the third region of the target DNA and the fifth DNA sequence is substantially non-homologous to the fourth region of the target DNA. The negative selection markers contained in such a PNS vector may either be the same or different negative selection markers. When they are different such that they require the use of two different agents to select again cells containing such negative markers, such negative selection may be carried out sequentially or simultaneously with appropriate agents for the negative selection marker. The positioning of two negative selection markers at the 5' and 3' end of a PNS vector further enhances selection against target cells which have randomly integrated the PNS vector. This is because random integration sometimes results in the rearrangement of the PNS vector resulting in excision of all or part of the negative selection marker prior to random integration. When this occurs, cells randomly integrating the PNS vector cannot be selected against. However, the presence of a second negative selection marker on the PNS vector substantially enhances the likelihood that random integration will result in the insertion of at least one of the two negative selection markers.

The substantial non-homology between the fourth DNA sequence (and in some cases fourth and fifth DNA sequences) of the PNS vector and the target DNA creates a discontinuity in sequence homology at or near the juncture of the fourth DNA sequence. Thus, when the vector is integrated into the genome by way of the homologous recombination mechanism of the cell, the negative selection marker in the fourth DNA sequence is not transferred into the target DNA. It is the non-integration of this negative selection marker during homologous recombination which forms the basis of the PNS method of the invention.

As used herein, a "modifying DNA sequence" is a DNA sequence contained in the first, second and/or third DNA sequence which encodes the substitution, insertion and/or deletion of one or more nucleotides in the target DNA sequence after homologous insertion of the PNS vector into the targeted region of the genome. When the PNS vector contains only the insertion of the third DNA sequence encoding the positive selection marker, the third DNA sequence is sometimes referred to as a "first modifying DNA sequence". When in addition to the third DNA sequence, the PNS vector also encodes the further substitution, insertion and/or deletion of one or more nucleotides, that portion encoding such further modification is sometimes referred to as a "second modifying DNA sequence". The second modifying DNA sequence may comprise the entire first and/or second DNA sequence or in some instances may comprise less than the entire first and/or second DNA sequence. The latter case typically arises when, for example, a heterologous gene is incorporated into a PNS vector which is designed to place that heterologous gene under the regulatory control of endogenous regulatory sequences. In such a case, the homologous portion of, for example, the first DNA sequence may comprise all or part of the targeted endogenous regulatory sequence and the modifying DNA sequence comprises that portion of the first DNA sequence (and in some cases a part of the second DNA sequence as well) which encodes the heterologous DNA sequence. An appropriate homologous portion in the second DNA sequence will be included to complete the targeting of the PNS vector. On the other hand, the entire first and/or second DNA sequence may comprise a second modifying DNA sequence when, for example, either or both of these DNA sequences encode for the correction of a genetic defect in the targeted DNA sequence.

As used herein, "modified target DNA sequence" refers to a DNA sequence in the genome of a targeted cell which has been modified by a PNS vector. Modified DNA sequences contain the substitution, insertion and/or deletion of one or more nucleotides in a first transformed target cell as compared to the cells from which such transformed target cells are derived. In some cases, modified target DNA sequences are referred to as "first" and/or "second modified target DNA sequences". These correspond to the DNA sequence found in the transformed target cell when a PNS vector containing a first or second modifying sequence is homologously integrated into the target DNA sequence.

"Transformed target cells" sometimes referred to as "first transformed target cells" refers to those target cells wherein the PNS vector has been homologously integrated into the target cell genome. A "transformed cell" on the other hand refers to a cell wherein the PNS has non-homologously inserted into the genome randomly. "Transformed target cells" generally contain a positive selection marker within the modified target DNA sequence. When the object of the genomic modification is to disrupt the expression of a particular gene, the positive selection marker is generally contained within an exon which effectively disrupts transcription and/or translation of the targeted endogenous gene. When, however, the object of the genomic modification is to

insert an exogenous gene or correct an endogenous gene defect, the modified target DNA sequence in the first transformed target cell will in addition contain exogenous DNA sequences or endogenous DNA sequences corresponding to those found in the normal, i.e., nondefective, endogenous gene.

"Second transformed target cells" refers to first transformed target cells whose genome has been subsequently modified in a predetermined way. For example, the positive selection marker contained in the genome of a first transformed target cell can be excised by homologous recombination to produce a second transformed target cell. The details of such a predetermined genomic manipulation will be described in more detail hereinafter.

As used herein, "heterologous DNA" refers to a DNA sequence which is different from that sequence comprising the target DNA sequence. Heterologous DNA differs from target DNA by the substitution, insertion and/or deletion of one or more nucleotides. Thus, an endogenous gene sequence may be incorporated into a PNS vector to target its insertion into a different regulatory region of the genome of the same organism. The modified DNA sequence so obtained is a heterologous DNA sequence. Heterologous DNA sequences also include endogenous sequences which have been modified to correct or introduce gene defects or to change the amino acid sequence encoded by the endogenous gene. Further, heterologous DNA sequences include exogenous DNA sequences which are not related to endogenous sequences, e.g. sequences derived from a different species. Such "exogenous DNA sequences" include those which encode exogenous polypeptides or exogenous regulatory sequences. For example, exogenous DNA sequences which can be introduced into murine or bovine ES cells for tissue specific expression (e.g. in mammary secretory cells) include human blood factors such as t-PA, Factor VIII, serum albumin and the like. DNA sequences encoding positive selection markers are further examples of heterologous DNA sequences.

The PNS vector is used in the PNS method to select for transformed target cells containing the positive selection marker and against those transformed cells containing the negative selection marker. Such positive-negative selection procedures substantially enrich for those transformed target cells wherein homologous recombination has occurred. As used herein, "substantial enrichment" refers to at least a two-fold enrichment of transformed target cells as compared to the ratio of homologous transformants versus nonhomologous transformants, preferably a 10-fold enrichment, more preferably a 1000-fold enrichment, most preferably a 10,000-fold enrichment, i.e., the ratio of transformed target cells to transformed cells. In some instances, the frequency of homologous recombination versus random integration is of the order of 1 in 1000 and in some cases as low as 1 in 10,000 transformed cells. The substantial enrichment obtained by the PNS vectors and methods of the invention often result in cell populations wherein about 1%, and more preferably about 20%, and most preferably about 95% of the resultant cell population contains transformed target cells wherein the PNS vector has been homologously integrated. Such substantially enriched transformed target cell populations may thereafter be used for subsequent genetic manipulation, for cell culture experiments or for the production of transgenic organisms such as transgenic animals or plants.

FIGS. 2*a* and 2*b* show the consequences of gene targeting (homologous recombination) and random integration of a PNS vector into the genome of a target cell. The PNS vector shown contains a neomycin resistance gene as a positive selection marker (neo^r) and a herpes simplex virus thymidine kinase (HSV-tk) gene as a negative selection marker. The neo^r positive selection marker is positioned in an exon of gene X. This positive selection marker is constructed such that it's expression is under the independent control of appropriate regulatory sequences. Such regulatory sequences may be endogenous to the host cell in which case they are preferably derived from genes actively expressed in the cell type. Alteratively, such regulatory sequences may be inducible to permit selective activation of expression of the positive selection marker.

On each side of the neo^r marker are DNA sequences homologous to the regions 5' and 3' from the point of neo^r insertion in the exon sequence. These flanking homologous sequences target the X gene for homologous recombination with the PNS vector. Consistent with the above description of the PNS vector, the negative selection marker HSV-tk is situated outside one of the regions of homology. In this example it is 3' to the transcribed region of gene X. The neo^r gene confers resistance to the drug G418 (G418^R) whereas the presence of the HSV-tk gene renders cells containing this gene sensitive to gancyclovir (GANC^S). When the PNS vector is randomly inserted into the genome by a mechanism other than by homologous recombination (FIG. 2*b*), insertion is most frequently via the ends of the linear DNA and thus the phenotype for such cells is neo^r HSV-tk^r (G418^R, GANC^S). When the PNS vector is incorporated into the genome by homologous recombination as in FIG. 2*a*, the resultant phenotype is neo^r, HSV-tk$^-$(G418^R, GANC^R). Thus, those cells wherein random integration of the PNS vector has occurred can be selected against by treatment with GANC. Those remaining transformed target cells wherein homologous recombination has been successful can then be selected on the basis of neomycin resistance and GANC resistance. It, of course, should be apparent that the order of selection for and selection against a particular genotype is not important and that in some instances positive and negative selection can occur simultaneously.

As indicated, the neomycin resistance gene in FIG. 2 is incorporated into an exon of gene X. As so constructed, the integration of the PNS vector by way of homologous recombination effectively blocks the expression of gene X. In multicellular organisms, however, integration is predominantly random and occurs, for the most part, outside of the region of the genome encoding gene X. Non-homologous recombination therefore will not disrupt gene X in most instances. The resultant phenotypes will therefore, in addition to the foregoing, will also be X$^-$ for homologous recombination and X$^+$ for random integration. In many cases it is desirable to disrupt genes by positioning the positive selection marker in an exon of a gene to be disrupted or modified. For example, specific proto-oncogenes can be mutated by this method to produce transgenic animals. Such transgenic animals containing selectively inactivated proto-oncogenes are useful in dissecting the genetic contribution of such a gene to oncogenesis and in some cases normal development.

Another potential use for gene inactivation is disruption of proteinaceous receptors on cell surfaces. For example, cell lines or organisms wherein the expression of a putative viral receptor has been disrupted using an appropriate PNS vector can be assayed with virus to confirm that the receptor is, in fact, involved in viral infection. Further, appropriate PNS vectors may be used to produce transgenic animal models for specific genetic defects. For example, many gene defects have been characterized by the failure of specific genes to express functional gene product, e.g. α and β

thalassema, hemophilia, Gaucher's disease and defects affecting the production of α-1-antitrypsin, ADA, PNP, phenylketonurea, familial hypercholesterolemia and retinoblastemia. Transgenic animals containing disruption of one or both alleles associated with such disease states or modification to encode the specific gene defect can be used as models for therapy. For those animals which are viable at birth, experimental therapy can be applied. When, however, the gene defect affects survival, an appropriate generation (e.g. F0, F1) of transgenic animal may be used to study in vivo techniques for gene therapy.

A modification of the foregoing means to disrupt gene X by way of homologous integration involves the use of a positive selection marker which is deficient in one or more regulatory sequences necessary for expression. The PNS vector is constructed so that part but not all of the regulatory sequences for gene X are contained in the PNS vector 5' from the structural gene segment encoding the positive selection marker, e.g., homologous sequences encoding part of the promotor of the X gene. As a consequence of this construction, the positive selection marker is not functional in the target cell until such time as it is homologously integrated into the promotor region of gene X. When so integrated, gene X is disrupted and such cells may be selected by way of the positive selection marker expressed under the control of the target gene promoter. The only limitation in using such an approach is the requirement that the targeted gene be actively expressed in the cell type used. Otherwise, the positive selection marker will not be expressed to confer a positive selection characteristic on the cell.

In many instances, the disruption of an endogenous gene is undesirable, e.g., for some gene therapy applications. In such situations, the positive selection marker comprising the third DNA sequence of the PNS vector may be positioned within an untranslated sequence, e.g. an intron of the target DNA or 5' or 3' untranslated regions. FIG. 3 depicts such a PNS vector. As indicated, the first DNA sequence comprises part of exon I and a portion of a contiguous intron in the target DNA. The second DNA sequence encodes an adjacent portion of the same intron and optionally may include all or a portion of exon II. The positive selection marker of the third DNA sequence is positioned between the first and second sequences. The fourth DNA sequence encoding the negative selection marker, of course, is positioned outside of the region of homology. When the PNS vector is integrated into the target DNA by way of homologous recombination the positive selection marker is located in the intron of the targeted gene. The third DNA sequence is constructed such that it is capable of being expressed and translated independently of the targeted gene. Thus, it contains an independent functional promotor, translation initiation sequence, translation termination sequence, and in some cases a polyadenylation sequence and/or one or more enhancer sequences, each functional in the cell type transfected with the PNS vector. In this manner, cells incorporating the PNS vector by way of homologous recombination can be selected by way of the positive selection marker without disruption of the endogenous gene. Of course, the same regulatory sequences can be used to control the expression of the positive selection marker when it is positioned within an exon. Further, such regulatory sequences can be used to control expression of the negative selection marker. Regulatory sequences useful in controlling the expression of positive and/or negative selection markers are listed in Table IIB. Of course, other regulatory sequences may be used which are known to those skilled in the art. In each case, the regulatory sequences will be properly aligned and, if necessary, placed in proper reading frame with the particular DNA sequence to be expressed. Regulatory sequence, e.g. enhancers and promoters from different sources may be combined to provide modulated gene expression.

TABLE IIA

Tissue Specific Regulatory Sequences		
Cell/Tissue	Promoter/Enhancer	Reference
Adrenal	PNMT	Bartge, et al. (1988) PNAS 85
Erythroid	β-globin	Townes et al. (1985) EMBO J 4:1715
Lens	α-crystallin	Overteek et al. (1985) PNAS 82:7815
Liver	α-FP	Krumlauf et al. (1985) MCB 5:1639
Lymphoid	Igμ (γ-1) promoter/enhancer	Yamamura et al. (1986) PNAS 83:2152
Mammary	WAP	Gordon et al. (1987) Bio/Tech 5:1183
Nervous	MBP	Tamura et al. (1989) MCB 9:3122
Pancreas (B)	Insulin	Hanahan (1985) Nature 315:115
Pancreas (exocrine)	Elastase	Swift et al. (1984) Cell 38:639
Pituitary	Prolactin	Ingraham et al. (1988) Cell 55:579
Skeletal Muscle	ckm	Johnson et al. (1989) MCB 9:3393
Testes	Protamine	Stewart et al. (1988) MCB 8:1748

TABLE IIB

Regulatory Sequences for Use With Positive and/or Negative Selection Markers	
Regulatory Sequence	Cell Type
PYF441 enhancer/HSV-tk promoter (pMCI-Neo control)	embryo-derived
ASV-LTR	fibroblasts
SV-40 early	variety of mammalian cells
Cytomegalo virus	general mammalian
β-actin	general mammalian
MoMuLV	haemopoetic stem cells
SFFV	haemopoetic stem cells
Mannopine synthase	general plant
Octapine synthase	general plant
Nopaline synthase	general plant
Cauliflower mosiac virus 35S promotor/enhancer	general plant
β-phaseolin "insert-7"	seeds protoplasts

A modification of the target DNA sequence is also shown in FIG. 3. In exon I of the target DNA sequence, the sixth codon GTG is shown which encodes valine. In the first DNA sequence of the PNS vector, the codon GAG replaces the GTG codon in exon I. This latter codon encodes glutamine. Cells selected for homologous recombination as a consequence encode a modified protein wherein the amino acid encoded by the sixth codon is changed from valine to glutamine.

There are, of course, numerous other examples of modifications of target DNA sequences in the genome of the cell which can be obtained by the PNS vectors and methods of the invention. For example, endogenous regulatory sequences controlling the expression of proto-oncogenes can be replaced with regulatory sequences such as promoters and/or enhancers which actively express a particular gene in

a specific cell type in an organism, i.e., tissue-specific regulatory sequences. In this manner, the expression of a proto-oncogene in a particular cell type, for example in a transgenic animal, can be controlled to determine the effect of oncogene expression in a cell type which does not normally express the proto-oncogene. Alternatively, known viral oncogenes can be inserted into specific sites of the target genome to bring about tissue-specific expression of the viral oncogene. Examples of preferred tissue-specific regulatory sequences are listed in Table IIA. Examples of proto-oncogenes which may be modified by the PNS vectors and methods to produce tissue specific expression and viral oncogenes which may be placed under control of endogenous regulatory sequences are listed in Table IIIA and IIIB, respectively.

TABLE IIIA

Proto-oncogenes involved in human tumors

Gene	Disease
c-abl	chronic myelogenous leukemia
c-erbB	squamous cell carcinoma
	glial blastoma
c-myc	Burkitt's lymphoma
	small cell carcinoma of lung
	carcinoma of breat
L-myc	small cell carcinoma of lung
N-myc	small cell carcinoma of lung
	neuroblastoma
neu	carcinoma of breast
C-ras	variety

TABLE IIIB

Viral oncogenes known to cause tumors when ectopically expressed in mice

Ha-ras	Sv40Tag
HPV-16	v-abl
HPV-E7	v-fps
PyTag	v-myc
	v-src

As indicated, the positive-negative selection methods and vectors of the invention are used to modify target DNA sequences in the genome of target cells capable of homologous recombination. Accordingly, the invention may be practiced with any cell type which is capable of homologous recombination. Examples of such target cells include cells derived from vertebrates including mammals such as humans, bovine species, ovine species, murine species, simian species, and other eucaryotic organisms such as filamentous fungi, and higher multicellular organisms such as plants. The invention may also be practiced with lower organisms such as gram positive and gram negative bacteria capable of homologous recombination. However, such lower organisms are not preferred because they generally do not demonstrate significant non-homologous recombination, i.e., random integration. Accordingly, there is little or no need to select against non-homologous transformants.

In those cases where the ultimate goal is the production of a non-human transgenic animal, embryonic stem cells (ES cells) are preferred target cells. Such cells have been manipulated to introduce transgenes. ES cells are obtained from pre-implantation embryos cultured in vitro. Evans, M. J., et al. (1981), *Nature,* 292, 154–156; Bradley, M. O., et al. (1984), *Nature,* 309, 255–258; Gossler, et al. (1986), *Proc. Natl. Acad. Sci. U.S.A.*, 83, 9065–9069; and Robertson, et al. (1986), *Nature,* 322, 445–448. PNS vectors can be efficiently introduced into the ES cells by electroporation or microinjection or other transformation methods, preferably electroporation. Such transformed ES cells can thereafter be combined with blastocysts from a non-human animal. The ES cells thereafter colonize the embryo and can contribute to the germ line of the resulting chimeric animal. For review see Jaenisch, R. (1988), *Science,* 240, 1468–1474. In the present invention, PNS vectors are targeted to a specific portion of the ES cell genome and thereafter used to generate chimeric transgenic animals by standard techniques.

When the ultimate goal is gene therapy to correct a genetic defect in an organism such as a human being, the cell type will be determined by the etiology of the particular disease and how it is manifested. For example, hemopoietic stem cells are a preferred cells for correcting genetic defects in cell types which differentiate from such stem cells, e.g. erythrocytes and leukocytes. Thus, genetic defects in globin chain synthesis in erythrocytes such as sickle cell anemia, β-thalassemia and the like may be corrected by using the PNS vectors and methods of the invention with hematopoietic stem cells isolated from an affected patient. For example, if the target DNA in FIG. 3 is the sickle-cell β-globin gene contained in a hematopoietic stem cell and the PNS vector in FIG. 3 is targeted for this gene with the modification shown in the sixth codon, transformed hematopoietic stem cells can be obtained wherein a normal β-globin will be expressed upon differentiation. After correction of the defect, the hematopoietic stem cells may be returned to the bone marrow or systemic circulation of the patient to form a subpopulation of erythrocytes containing normal hemoglobin. Alternatively, hematopoietic stem cells may be destroyed in the patient by way of irradiation and/or chemotherapy prior to reintroduction of the modified hematopoietic stem cell thereby completely rectifying the defect.

Other types of stem cells may be used to correct the specific gene defects associated with cells derived from such stem cells. Such other stem cells include epithelial, liver, lung, muscle, endothelial, menchymal, neural and bone stem cells. Table IV identifies a number of known genetic defects which are amenable to correction by the PNS methods and vectors of the invention.

Alternatively, certain disease states can be treated by modifying the genome of cells in a way which does not correct a genetic defect per se but provides for the supplementation of the gene product of a defective gene. For example, endothelial cells are preferred as targets for human gene therapy to treat disorders affecting factors normally present in the systemic circulation. In model studies using both dogs and pigs endothelial cells have been shown to form primary cultures, to be transformable with DNA in culture, and to be capable of expressing a transgene upon re-implantation in arterial grafts into the host organism. Wilson, et al. (1989), *Science,* 244, 1344; Nabel, et al. (1989), *Science,* 244, 1342. Since endothelial cells form an integral part of the graft, such transformed cells can be used to produce proteins to be secreted into the circulatory system and thus serve as therapeutic agents in the treatment of genetic disorders affecting circulating factors. Examples of such diseases include insulin-deficient diabetes, α-1-antitrypsin deficiency, and hemophilia. Epithelial cells provide a particular advantage in the treatment of factor VIII-deficient hemophilia. These cells naturally produce von Willebrand factor and it has been shown that production of active factor VIII is dependant upon the autonomous synthesis of vWF (Toole, et al. (1986), *Proc. Natl. Acad. Sci. U.S.A.*, 83, 5939).

As indicated in Example 4, human endothelial cells from a hemophiliac patient deficient in Factor VIII are modified

by a PNS vector to produce an enriched population of transformed endothelial cells wherein the expression of DNA sequences encoding a secretory form of Factor VIII is placed under the control of the regulatory sequences of the endogenous β-actin gene. Such transformed cells are implanted into vascular grafts from the patient. After incorporation of transformed cells, it is grafted back into the vascular system of the patient. The transformed cells secrete Factor XIII into the vascular system to supplement the defect in the patients blood clotting system.

Other diseases of the immune and/or the circulatory system are candidates for human gene therapy. The target tissue, bone marrow, is readily accessible by current technology, and advances are being made in culturing stem cells in vitro. The immune deficiency diseases caused by mutations in the enzymes adenosine deaminase (ADA) and purine nucleotide phosphorylase (PNP), are of particular interest. Not only have the genes been cloned, but cells corrected by PNS gene therapy are likely to have a selective advantage over their mutant counterparts. Thus, ablation of the bone marrow in recipient patients may not be necessary.

The PNS approach is applicable to genetic disorders with the following characteristics: first, the DNA sequence and preferably the cloned normal gene must be available; second, the appropriate, tissue relevant, stem cell or other appropriate cell must be available. Below is Table IV listing some of the known genetic diseases, the name of the cloned gene, and the tissue type in which therapy may be appropriate. These and other genetic disease amenable to the PNS methods and vectors of the invention have been reviewed. See Friedman (1989), *Science*, 244, 1275; Nichols, E. K. (1988), *Human Gene Therapy* (Harvard University Press); and *Cold Springs Harbor Symposium on Quantitative Biology*, Vol. 11 (1986), "The Biology of Homo Sapiens" (Cold Springs Harbor Press).

TABLE IV

Human Genetic Diseases in Which the Disease Locus has been Cloned

Disease	Gene	Target Tissue
α1-anti-trypsin disease	α1-anti trypsin	liver
Gaucher Disease	glucocerebrosidase	bone marrow
Granulocyte Actin Deficiency	Granulocyte Actin	bone marrow
Immunodeficiency	Adenosine deaminase	bone marrow
Immunodeficiency	Purine nucleoside	bone marrow
Muscular Dystrophy	most likely dystropin gene	skeletal muscle
Phenylketonuria	Phenylalanine hydroxylase	liver
Sickle Cell Anemia	β-globin	bone marrow
Thalassemia	globin	bone marrow
Hemophilia	various clotting factors	bone marrow/ endothelial cells
Familial hyper-cholesterolemia	low density lipoprotein receptor	liver/endo-endothelial cells

As indicated, genetic defects may be corrected in specific cell lines by positioning the positive selection marker (the second DNA sequence in the PNS vector) in an untranslated region such as an intron near the site of the genetic defect together with flanking segments to correct the defect. In this approach, the positive selection marker is under its own regulatory control and is capable of expressing itself without substantially interfering with the expression of the targeted gene. In the case of human gene therapy, it may be desirable to introduce only those DNA sequences which are necessary to correct the particular genetic defect. In this regard, it is desirable, although not necessary, to remove the residual positive selection marker which remains after correction of the genetic defect by homologous recombination.

The removal of a positive selection marker from a genome in which homologous insertion of a PNS vector has occurred can be accomplished in many ways. For example, the PNS vector can include a second negative selection marker contained within the third DNA sequence. This second negative selection marker is different from the first negative selection marker contained in the fourth DNA sequence. After homologous integration, a second modified target DNA sequence is formed containing the third DNA sequence encoding both the positive selection marker and the second negative selection marker. After isolation and purification of the first transformed target cells by way of negative selection against transformed cells containing the first negative selection marker and for those cells containing the positive selection marker, the first transformed target cells are subjected to a second cycle of homologous recombination. In this second cycle, a second homologous vector is used which contains all or part of the first and second DNA sequence of the PNS vector (encoding the second modification in the target DNA) but not those sequences encoding the positive and second negative selection markers. The second negative selection marker in the first transformed target cells is then used to select against unsuccessful transformants and cells wherein the second homologous vector is randomly integrated into the genome. Homologous recombination of this second homologous vector, however, with the second modified target DNA sequence results in a second transformed target cell type which does not contain either the positive selection marker or the second negative selection marker but which retains the modification encoded by the first and/or second DNA sequences. Cells which have not homologously integrated the second homologous vector are selected against using the second negative selection marker.

The PNS vectors and methods of the invention are also applicable to the manipulation of plant cells and ultimately the genome of the entire plant. A wide variety of transgenic plants have been reported, including herbaceous dicots, woody dicots and monocots. For a summary, see Gasser, et al. (1989), *Science*, 244, 1293–1299. A number of different gene transfer techniques have been developed for producing such transgenic plants and transformed plant cells. One technique used *Agrobacterium tumefaciens* as a gene transfer system. Rogers, et al. (1986), *Methods Enzymol.*, 118, 627–640. A closely related transformation utilizes the bacterium *Agrobacterium rhizogenes*. In each of these systems a Ti or Ri plant transformation vector can be constructed containing border regions which define the DNA sequence to be inserted into the plant genome. These systems previously have been used to randomly integrate exogenous DNA to plant genomes. In the present invention, an appropriate PNS vector may be inserted into the plant transformation vector between the border sequences defining the DNA sequences transferred into the plant cell by the Agrobacterium transformation vector.

Preferably, the PNS vector of the invention is directly transferred to plant protoplasts by way of methods analogous to that previously used to introduce transgenes into protoplasts. See, e.g. Paszkowski, et al. (1984), *EMBO J.*, 3, 2717–2722; Hain, et al. (1985), *Mol. Gen. Genet.*, 199, 161–168; Shillito, et al. (1985), *Bio./Technology*, 3,

1099–1103; and Negrutiu, et al. (1987), *Plant Mol. Bio.*, 8, 363–373. Alternatively, the PNS vector is contained within a liposome which may be fused to a plant protoplast (see, e.g. Deshayes, et al. (1985), *EMBO J.*, 4, 2731–2738) or is directly inserted to plant protoplast by way of intranuclear microinjection (see, e.g. Crossway. et al. (1986), *Mol. Gen Genet.*, 202, 179–185, and Reich, et al. (1986), *Bio/Technology*, 4, 1001–1004). Microinjection is the preferred method for transfecting protoplasts. PNS vectors may also be microinjected into meristematic inflorenscences. De la Pena et al. (1987), *Nature*, 325, 274–276. Finally, tissue explants can be transfected by way of a high velocity microprojectile coated with the PNS vector analogous to the methods used for insertion of transgenes. See, e.g. Vasil (1988), *Bio/Technology*, 6, 397; Klein, et al. (1987), *Nature*, 327, 70; Klein, et al. (1988), *Proc. Natl. Acad. Sci. U.S.A.*, 85, 8502; McCabe, et al. (1988), *Bio/Technology*, 6, 923; and Klein, et al., *Genetic Engineering*, Vol 11, J. K. Setlow editor (Academic Press, N.Y., 1989). Such transformed explants can be used to regenerate for example various serial crops. Vasil (1988), *Bio/Technology*, 6, 397.

Once the PNS vector has been inserted into the plant cell by any of the foregoing methods, homologous recombination targets the PNS vector to the appropriate site in the plant genome. Depending upon the methodology used to transfect, positive-negative selection is performed on tissue cultures of the transformed protoplast or plant cell. In some instances, cells amenable to tissue culture may be excised from a transformed plant either from the F0 or a subsequent generation.

The PNS vectors and method of the invention are used to precisely modify the plant genome in a predetermined way. Thus, for example, herbicide, insect and disease resistance may be predictably engineered into a specific plant species to provide, for example, tissue specific resistance, e.g., insect resistance in leaf and bark. Alternatively, the expression levels of various components within a plant may be modified by substituting appropriate regulatory elements to change the fatty acid and/or oil content in seed, the starch content within the plant and the elimination of components contributing to undesirable flavors in food. Alternatively, heterologous genes may be introduced into plants under the predetermined regulatory control in the plant to produce various hydrocarbons including waxes and hydrocarbons used in the production of rubber.

The amino acid composition of various storage proteins in wheat and corn, for example, which are known to be deficient in lysine and tryptophan may also be modified. PNS vectors can be readily designed to alter specific codons within such storage proteins to encode lysine and/or tryptophan thereby increasing the nutritional value of such crops. For example, the zein protein in corn (Pederson et al. (1982), *Cell*, 29, 1015) may be modified to have a higher content of lysine and tryptophan by the vectors and methods of the invention.

It is also possible to modify the levels of expression of various positive and negative regulatory elements controlling the expression of particular proteins in various cells and organisms. Thus, the expression level of negative regulatory elements may be decreased by use of an appropriate promotor to enhance the expression of a particular protein or proteins under control of such a negative regulatory element. Alternatively, the expression level of a positive regulatory protein may be increased to enhance expression of the regulated protein or decreased to reduce the amount of regulated protein in the cell or organism.

The basic elements of the PNS vectors of the invention have already been described. The selection of each of the DNA sequences comprising the PNS vector, however, will depend upon the cell type used, the target DNA sequence to be modified and the type of modification which is desired.

Preferably, the PNS vector is a linear double stranded DNA sequence. However, circular closed PNS vectors may also be used. Linear vectors are preferred since they enhance the frequency of homologous integration into the target DNA sequence. Thomas, et al. (1986), *Cell*, 44, 49.

In general, the PNS vector (including first, second, third and fourth DNA sequences) has a total length of between 2.5 kb (2500 base pairs) and 1000 kb. The lower size limit is set by two criteria. The first of these is the minimum necessary length of homology between the first and second sequences of the PNS vector and the target locus. This minimum is approximately 500 bp (DNA sequence 1 plus DNA sequence 2). The second criterion is the need for functional genes in the third and fourth DNA sequences of the PNS vector. For practical reasons, this lower limit is approximately 1000 bp for each sequence. This is because the smallest DNA sequences encoding known positive and negative selection markers are about 1.0–1.5 kb in length.

The upper limit to the length of the PNS vector is determined by the state of the technology used to manipulate DNA fragments. If these fragments are propagated as bacterial plasmids, a practical upper length limit is about 25 kb; if propagated as cosmids, the limit is about 50 kb, if propagated as YACs (yeast artificial chromosomes) the limit approaches 1000 kb (Burke, et al. (1987), *Science*, 236, 806).

Within the first and second DNA sequences of the PNS vector are portions of DNA sequence which are substantially homologous with sequence portions contained within the first and second regions of the target DNA sequence. The degree of homology between the vector and target sequences influences the frequency of homologous recombination between the two sequences. One hundred percent sequence homology is most preferred, however, lower sequence homology can be used to practice the invention. Thus, sequence homology as low as about 80% can be used. A practical lower limit to sequence homology can be defined functionally as that amount of homology which if further reduced does not mediate homologous integration of the PNS vector into the genome. Although as few as 25 bp of 100% homology are required for homologous recombination in mammalian cells (Ayares, et al. (1986), *Genetics*, 83, 5199–5203), longer regions are preferred, e.g., 500 bp, more preferably, 5000 bp, and most preferably, 25000 bp for each homologous portion. These numbers define the limits of the individual lengths of the first and second sequences. Preferably, the homologous portions of the PNS vector will be 100% homologous to the target DNA sequence, as increasing the amount of non-homology will result in a corresponding decrease in the frequency of gene targeting. If non-homology does exist between the homologous portion of the PNS vector and the appropriate region of the target DNA, it is preferred that the non-homology not be spread throughout the homologous portion but rather in discrete areas of the homologous portion. It is also preferred that the homologous portion of the PNS vector adjacent to the negative selection marker (fourth or fifth DNA sequence) be 100% homologous to the corresponding region in the target DNA. This is to ensure maximum discontinuity between homologous and non-homologous sequences in the PNS vector.

Increased frequencies of homologous recombination have been observed when the absolute amount of DNA sequence in the combined homologous portions of the first and second DNA sequence are increased. FIG. 4 depicts the targeting

frequency of the Hprt locus as a function of the extent of homology between an appropriate PNS vector and the endogenous target. A series of replacement (▲) and insertion (●) Hprt vectors were constructed that varied in the extent of homology to the endogenous Hprt gene. Hprt sequences in each vector were interrupted in the eighth exon with the neomycin resistance gene. The amount of Hprt sequence 3' to the neogene was kept constant to the amount of Hprt sequence 5' to the neo was varied. The absolute frequency of independent targeting events per total ES cells electroporated is plotted in FIG. 4 on the logarithmic scale as a function of the number of kilobases of Hprt sequence contained within the PNS vectors. See Capecchi, M. R. (1989), *Science*, 244, 1288–1292.

As previously indicated, the fourth DNA sequence containing the negative selection marker should have sufficient non-homology to the target DNA sequence to prevent homologous recombination between the fourth DNA sequence and the target DNA. This is generally not a problem since it is unlikely that the negative selection marker chosen will have any substantial homology to the target DNA sequence. In any event, the sequence homology between the fourth DNA sequence and the target DNA sequence should be less than about 50%, most preferably less than about 30%.

A preliminary assay for sufficient sequence non-homology between the fourth DNA sequence and the target DNA sequence utilizes standard hybridization techniques. For example, the particular negative selection marker may be appropriately labeled with a radioisotope or other detectable marker and used as a probe in a Southern blot analysis of the genomic DNA of the target cell. If little or no signal is detected under intermediate stringency conditions such as 3XSSC when hybridized at about 55° C., that negative selection marker should be functional in a PNS vector designed for homologous recombination in that cell type. However, even if a signal is detected, it is not necessarily indicative that particular negative selection cannot be used in a PNS vector targeted for that genome. This is because the negative selection marker may be hybridizing with a region of the genome which is not in proximity with the target DNA sequence. Since the target DNA sequence is defined as those DNA sequences corresponding to first, second, third, and in some cases, fourth regions of the genome, Southern blots localizing the regions of the target DNA sequence may be performed. If the probe corresponding to the particular negative selection marker does not hybridize to these bands, it should be functional for PNS vectors directed to these regions of the genome.

Hybridization between sequences encoding the negative selection marker and the genome or target regions of a genome, however, does not necessarily mean that such a negative selection marker will not function in a PNS vector. The hybridization assay is designed to detect those sequences which should function in the PNS vector because of their failure to hybridize to the target. Ultimately, a DNA sequence encoding a negative selection marker is functional in a PNS vector if it is not integrated during homologous recombination regardless of whether or not it hybridizes with the target DNA.

It is also possible that high stringency hybridization can be used to ascertain whether genes from one species can be targeted into related genes in a different species. For example, preliminary gene therapy experiments may require that human genomic sequences replace the corresponding related genomic sequence in mouse cells. High stringency hybridization conditions such as 0.1 XSSC at about 68° C. can be used to correlate hybridization signal under such conditions with the ability of such sequences to act as homologous portions in the first and second DNA sequence of the PNS vector. Such experiments can be routinely performed with various genomic sequences having known differences in homology. The measure of hybridization may therefore correlate with the ability of such sequences to bring about acceptable frequencies of recombination.

Table I identifies various positive and negative selection markers which may be used respectively in the third and fourth DNA sequences of the PNS vector together with the conditions used to select for or against cells expressing each of the selection markers. As for animal cells such as mouse L cells, ES cells, preferred positive selection markers include DNA sequences encoding neomycin resistance and hygromycin resistance, most preferably neomycin resistance. For plant cells preferred positive selection markers include neomycin resistance and bleomycin resistance, most preferably neomycin resistance.

For animal cells, preferred negative selection markers include gpt and HSV-tk, most preferably HSV-tk. For plant cells, preferred negative selection markers include Gpt and HSV-tk. As genes responsible for bacterial and fungal pathogenesis in plants are cloned, other negative markers will become readily available.

As used herein, a "positive screening marker" refers to a DNA sequence used in a phage rescue screening method to detect homologous recombination. An example of such a positive screening marker is the supF gene which encodes a tyrosine transfer RNA which is capable of suppressing amber mutations. See Smithies, et al. (1985), *Nature*, 317, 230–234.

The following is presented by way of example and is not to be construed as a limitation on the scope of the invention.

EXAMPLE 1

Inactivation at the int-2 locus in mouse ES cells

1. PNS Vector Construction

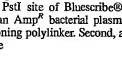
The PNS vector, pINT-2-*N*TK, is described in Mansour, et al. (1988), *Nature*, 336, 349. This vector was used to disrupt the proto-oncogene, INT-2, in mouse ES cells. As shown in FIG. 5*c*, it contains DNA sequences 1 and 2 homologous to the target INT-2 genomic sequences in mouse ES cells. These homologous sequences were obtained from a plasmid referred to as pAT-153 (Peters, et al. (1983), *Cell*, 33, 369). DNA sequence 3, the positive selection moiety of the PNS vector was the Neogene from the plasmid pMCINeo described in Thomas, et al. (1987), *Cell*, 51, 503; DNA sequence 4, the negative selection element of the vector, was the HSV-TK gene derived from the plasmid pIC-19-R/TK which is widely available in the scientific community.

Plasmid pIC19R/MC1-TK (FIG. 5*d*) contains the HSV-TK gene engineered for expression in ES cells (Mansour, et al. (1988), *Nature*, 336, 348–352). The TK gene, flanked by a duplication of a mutant polyoma virus enhancer, PYF441, has been inserted into the vector, pIC19R (Marsh, et al. (1984), *Gene*, 32, 481–485) between the XhoI and the HindIII sites. The map of plasmid pIC19R/MC1-TK is shown in FIG. 5*d*. The enhancer sequence is as follows:

5'
CTCGAGCAGTGTGGTTTTCAAGAGGAAGCAAAAAGCCTCTCCACCCAGGC
CTGGAATGTTTCCACACCCAATGTCGAGCAGTGTGGTTTTCAAGAGGAAGC
AAAAAGCCTC TCCACCCAGG CCTGGAATGT TTCCACCCAA TGTCGAG
3'

The 5' end is an XhoI restriction enzyme site, the 3' end is contiguous with the HSV-TK gene. The HSV-TK sequences are from nucleotides 92–1799 (McKnight (1980), *Nucl. Acids. Res.*, 8, 5949–5964) followed at the 3' end by a HindIII linker. The plasmid pIC19R is essentially identical to the pUC vectors, with an alternative poly-linker as shown in FIG. 5*d*.

Construction of the vector, pINT-2-N/TK involved five sequential steps as depicted in FIG. 5. First, a 3,965 bp PstI fragment containing exon 1*b*, was excised from pAT153 and inserted into the PstI site of Bluescribe© (Stratagene of LaJolla, Calif.), an Amp^R bacterial plasmid containing a multi-enzyme, cloning polylinker. Second, a synthetic XhoI linker of sequence



was inserted into the ApaI site on exon 1*b*. Third, the XhoI-SalI Neo^r-fragment from pMCI Neo was inserted into the XhoI linker in exon 1*b*. Fourth, the 3,965 bp INT-2 Pst fragment containing the Neo^r gene was reinserted into pAT153, to generate the plasmid pINT-2-N as shown in FIG. 5*b*. This plasmid also includes the third exon of the int-2 gene. Fifth, the ClaI-HindII HSV-tk fragment from pIC-19-R/TK was inserted into ClaI-HindII digested pINT2-N, creating the final product, pINT2-N/TK. This vector was linearized by digestion with ClaI prior to its introduction into ES cells.

2. Generation of ES Cells

ES cells were derived from two sources. The first source was isolation directly from C57B½ blastocysts (Evans, et al. (1981), *Nature*, 292, 154–156) except that primary embryonic fibroblasts (Doetschman, et al. (1985), *J. Embryol. Exp. Morphol.*, 87, 27–45) were used as feeders rather than STO cells. Briefly, 2.5 days postpregnancy mice were ovariectomized, and delayed blastocysts were recovered 4–6 days later. The blastocysts were cultured on mitomycin C-inactivated primary embryonic fibroblasts. After blastocyst attachment and the outgrowth of the trophectoderm, the ICM-derived clump was picked and dispersed by trypsin into clumps of 3–4 cells and put onto new feeders. All culturing was carried out in DMEM plus 20% FCS and 10^{-4}M β-mercaptoethanol. The cultures were examined daily. After 6–7 days in culture, colonies that still resembled ES cells were picked, dispersed into single cells, and replated on feeders. Those cell lines that retained the morphology and growth characteristic of ES cells were tested for pluripotency in vitro. These cell lines were maintained on feeders and transferred every 2–3 days.

The second method was to utilize one of a number of ES cell lines isolated from other laboratories, e.g., CC1.2 described by Kuehn, et al. (1987), *Nature*, 326, 295. The cells were grown on mitomycin C-inactivated STO cells. Cells from both sources behaved identically in gene targeting experiments.

3. Introduction of PNS Vector pINT-2-N/TK into ES cells

The PNS vector pINT-2-N/TK was introduced into ES cells by electroporation using the Promega Biotech X-Cell 2000. Rapidly growing cells were trypsinized, washed in DMEM, counted and resuspended in buffer containing 20 mM HEPES (pH 7.0), 137 mM NaCl, 5 mM KCl, 0.7 mM Na_2HPO_4, 6 mM dextrose, and 0.1 mM β-mercaptoethanol. Just prior to electroporation, the linearized recombinant vector was added. Approximately 25 μg of linearized PNS vector was mixed with 10^7 ES cells in each 1 ml-cuvette.

Cells and DNA were exposed to two sequential 625 V/cm pulses at room temperature, allowed to remain in the buffer for 10 minutes, then plated in non-selective media onto feeder cells.

4. Selection of ES Cells Containing a Targeted Disruption of the int-2 Locus

Following two days of non-selective growth, the cells were trypsinized and replated onto G418 (250 μg/ml) media. The positive-selection was applied alone for three days, at which time the cells were again trypsinized and replated in the presence of G418 and either ganycylovir (2×10^{-6}M) (Syntex, Palo Alto, Calif.) or 1-(2-deoxy-2-fluoro-β-D-arabino-furanosyl- 5-iodouracil (FIAU) (1×10^{-6}M) (Bristol Myers). When the cells had grown to confluency, each plate of cells was divided into two aliquots, one of which was frozen in liquid N_2, the other harvested for DNA analysis.

5. Formation of INT-2 disrupted transgenic mice

Those transformed cells determined to be appropriately modified by the PNS vector were grown in non-selective media for 2–5 days prior to injection into blastocysts according to the method of Bradley in *Teratocarcinomas and embryonic stem cells, a practical approach*, edited by E. J. Robertson, IRL Press, Oxford (1987), p. 125.

Blastocysts containing the targeted ES cells were implanted into pseudo-pregnant females and allowed to develop to term. Chimaeric offspring were identified by coat-color markers and those males showing chimaerism were selected for breeding offspring. Those offspring which carry the mutant allele can be identified by coat color, and the presence of the mutant allele reaffirmed by DNA analysis by tail-blot, DNA analysis.

EXAMPLE 2

Disruption at the hoxl.4 locus in mouse ES cells

Disruption of the hoxl.4 locus was performed by methods similar to those described to disrupt the int-2 locus. There were two major differences between these two disruption strategies. First, the PNS vector, pHOX1.4N/TK-TK2 (FIG. 6), used to disrupt the hoxl.4 locus contained two negative selection markers, i.e., a DNA sequence 5 encoding a second negative selection marker was included on the PNS vector at the end opposite to DNA sequence 4 encoding the first negative selection marker. DNA sequence 5 contained the tk gene isolated from HSV-type 2. It functioned as a negative-selectable marker by the same method as the original HSV-tk gene, but the two tk genes are 20% non-homologous. This non-homology further inhibits recombination between DNA sequences 4 and 5 in the vector which might

have inhibited gene-targeting. The second difference between the int-2 and the hoxl.4 disruption strategies is that the vector pHOX1.4N/TK-TK2 contains a deletion of 1000 bp of hoxl.4 sequences internal to the gene, i.e., DNA sequences 1 and 2 are not contiguous.

The HSV-tk2 sequences used in this construction were obtained from pDG504 (Swain, M. A. et al. (1983), *J. Virol.*, 46, 1045). The structural TK gene from pDG504 was inserted adjacent to the same promoter/enhancer sequences used to express both the Neo and HSV-tk genes, to generate the plasmid pIC20H/TK2.

Construction of pHOX1.4N/TK-TK2 proceeded in five sequential steps as depicted in FIG. 6. First a clone containing hoxl.4 sequences was isolated from a genomic λ library. The λ library was constructed by inserting EcoRI partially digested mouse DNA into the λ-DASH® (Stratagene) cloning phage. The hoxl.4 containing phage were identified by virtue of their homology to a synthetic oligonucleotide synthesized from the published sequence of the hoxl.4 locus. Tournier-Lasserve, et al. (1989), *Mol. Cell Biol.*, 9, 2273. Second, a 9 kb SalI-SpeI fragment containing the hoxl.4 homeodomain was inserted into Bluescribe®. Third, a 1 kb BglII fragment within the hoxl.4 locus was replaced with the Neo^r gene isolated from pMCI Neo, creating the plasmid pHOX1.4N. Fourth, the XhoI-SalI fragment by HSV-tk from pIC19R/TK was inserted into the SalI site of pHOX1.4N, generating the plasmid pHOX1.4N/TK. Fifth, the SalI-SpeI fragment from pHOX1.4N/TK was inserted into a SalI-XbaI digest of the plasmid pIC20HTK2, generating the final product, pHOX1.4N/TK/TK2. This vector was digested with SalI to form a linear PNS vector which was transfected into mouse ES cells as described in Example 1. Positive-negative selection and the method of forming transgenic mice was also as described in Example 1. Southern blots of somatic cells demonstrate that the disrupted hoxl.4 gene was transferred to transgenic offspring.

EXAMPLE 3

Inactivation of Other Hox Genes

The methods described in Examples 1 and 2 have also been used to disrupt the hoxl.3, hoxl.6, hox2.3, and int-1 loci in ES cells. The genomic sequences for each of these loci (isolated from the same -Dash library containing the hoxl.4 clone) were used to construct PNS vectors to target disruption of these genes. All of these PNS vectors contain the Neo-gene from pMCI-Neo as the positive selection marker and the HSV-tk and HSV-tk2 sequences as negative selection markers.

TABLE V

Other Murine Developmental Genes Inactivated by PNS			
Locus	Genomic Fragment	Sequence Ref.	Neo-Insertion Site
hox1.3	1kb Xba-HindIII	Tournier-lasserve, et al. (1989), MCE, 9, 2273	EcoRI-site in homeo-domain
hox1.6	13kb partial RI EMBO, 6, 2977	Baron, et al. (1987), homeo-domain	BglII-site in
hox2.3	12kb BamHI Genomics, 1, 182	Hart, et al. (1987), homeo-domain	BglII-site in
int-1	13kb BglII	van Ooyen et al. (1984), Cell, 39, 233	XhoI-site in exon 2

EXAMPLE 4

Vascular Graft Supplementing Factor VIII

In this example, a functional factor VIII gene is targeted by a PNS vector to the β-actin locus in human endothelial cells. When so incorporated, the expression of factor VIII is controlled by the β-actin promoter, a promoter known to function in nearly all somatic cells, including fibroblasts, epithelial and endothelial cells. PNS vector construction is as follows: In step IA (FIG. 7A), the 13.8 kb EcoRI fragment containing the entire human β-actin gene from the A-phage, 14TB (Leavitte, et al. (1984), *Mol. Cell Bio.*, 4, 1961) is inserted, using synthetic EcoRI/XhoI adaptors, into the XhoI site of the TK vector, pIC-19-R/TK to form plasmid pBact/TK. See FIG. 7A.

In step 1B (FIG. 7B), the 7.2 kb SalI fragment from a factor VIII cDNA clone including its native signal sequence (Kaufman, et al. (1988), *JBC*, 263, 6352; Toole, et al. (1986), *Proc. Natl. Acad. Sci. U.S.A.*, 83, 5939) is inserted next to the Neo^r gene in a pMCI derivative plasmid. This places the neo^r gene (containing its own promoter/enhancer) 3' to the polyadenylation site of factor VIII. This plasmid is designated pFVIII/Neo.

In step 2 (FIG. 7C), the factor VIII/Neo fragment is excised with XhoI as a single piece and inserted using synthetic XhoI/NcoI adaptors at the NcoI site encompassing the met-initiation codon in pBact/TK. This codon lies in the 2nd exon of the β-actin gene, well away from the promoter, such that transcription and splicing of the mRNA is in the normal fashion. The vector so formed is designated pBact/FVIII/Neo/TK.

This vector is digested with either ClaI or HindIII which acts in the polylinker adjacent to the TK gene. The linker vector is then introduced by electroporation into endothelial cells isolated from a hemophiliac patient. The cells are then selected for G418 and gancyclovir resistance. Those cells shown by DNA analysis to contain the factor VIII gene targeted to the β-actin locus or cells shown to express FVIII are then seeded into a vascular graft which is subsequently implanted into the patient's vascular system.

EXAMPLE 5

Replacement of a mutant PNP gene in human bone marrow stem cells using PNS

The genomic clone of a normal purine nucleoside phosphonylase (PNP) gene, available as a 12.4 kb, Xba-partial fragment (Williams, et al. (1984), *Nucl. Acids Res*, 12, 5779; Williams, et al. (1987), *J. Biol. Chem.*, 262, 2332) is inserted at the XbaI site in the vector, pIC-19-R/TK. The neo^r gene from pMCI-Neo is inserted, using synthetic BamHI/XhoI linkers, into the BamHI site in intron 1 of the PNP gene. The linearized version of this vector (cut with ClaI) is illustrated in FIG. 8.

Bone marrow stem cells from PNP patients transfected with this vector are selected for neo^r, gan^r, in culture, and those cells exhibiting replacement of the mutant gene with the vector gene are transplanted into the patient.

EXAMPLE 6

Inactivation by insertional mutagenesis of the Hox 1.1 locus in mouse ES cells, using a promoterless PNS vector

A promoterless positive selection marker is obtained using the Neo^R gene, excised at its 5' end by enzyme, EcoRI, from the plasmid, pMCI-Neo. Such a digestion removes the Neo structural gene from its controlling elements.

A promoterless PNS vector is used to insert the Neo gene into the Hox ½ gene in ES cells. The Hox 1.1 gene is expressed in cultured embryo cells (Colberg-Poley, et al. (1985), *Nature*, 314, 713) and the site of insertion, the second exon, lies 3' to the promoter of the gene (Kessel, et al. (1987), *PNAS*, 84, 5306; Zimmer, et al. (1989), *Nature*, 338, 150). Expression of Neo will thus be dependent upon insertion at the Hox 1.1 locus.

Vector construction is as follows:

Step 1—The neo gene, missing the transcriptional control sequences is removed from pMCI-Neo, and inserted into the second exon of the 11 kb, FspI-KpnI fragment of Hox 1.1 (Kessel, et al. (1987), supra; Zimmer, et al. (1989), supra).

Step 2—The Hox 1.1-Neo sequences is then inserted adjacent to the HSV-tk gene is pIC19R/TK, creating the targeting vector, pHox1.1-N/TK. The linearized version of this vector is shown in FIG. 9 This vector is electroporated into ES cells, which are then selected for Neo^r, GanC^r. The majority of cells surviving this selection are predicted to contain targeted insertions of Neo at the Hoxl.1 locus.

EXAMPLE 7

Inducible promoters

PNS vectors are used to insert novel control elements, for example inducible promoters, into specific genetic loci. This permits the induction of specified proteins under the spatial and/or temporal control of the investigator. In this example, the MT-1 promoter is inserted by PNS into the Int-2 gene in mouse ES cells.

The inducible promoter from the mouse metallothionein-I (MT-I) locus is targeted to the Int-2 locus. Mice generated from ES cells containing this alteration have an Int-2 gene inducible by the presence of heavy metals. The expression of this gene in mammary cells is predicted to result in oncogenesis and provides an opportunity to observe the induction of the disease.

Vector construction is as follows:

Step 1—The EcoRI-BglII fragment from the MT-I gene (Palmiter, et al. (1982), *Cell*, 29, 701) is inserted by blunt-end ligation into the BSSHII site, 5' to the Int-2 structural gene in the plasmid, pAT 153 (see discussion of Example 1).

Step 2—The MCI-Neo gene is inserted into the AvrII site in intron 2 of the Int-2-MT-I construct.

Step 3—The int-2-MT-IlNeo fragment is inserted into the vector, pIC 19R/TK, resulting in the construct shown in FIG. **10**.

Introduction of this gene into mouse ES cells by electroporation, followed by Neo^r, GanC^r, selection results in cells containing the MT-I promoter inserted 5' to the Int-2 gene. These cells are then inserted into mouse blastocysts to generate mice carrying this particular allele.

EXAMPLE 8

Inactivation of the ALS-II gene in tobacco protoplasts by PNS

A number of herbicides function by targeting specific plant metabolic enzymes. Mutant alleles of the genes encoding these enzymes have been identified which confer resistance to specific herbicides. Protoplasts containing these mutant alleles have been isolated in culture and grown to mature plants which retain the resistant phenotype (Botterman, et al. (1988), *TIGS*, 4, 219; Gasser, et al. (1989), *Science*, 244, 1293). One problem with this technology is that the enzymes involved are often active in multimer form, and are coded by more than one genetic locus. Thus, plants containing a normal (sensitive) allele at one locus and a resistant allele at another locus produce enzymes with mixed subunits which show unpredictable resistance characteristics.

In this example, the gene product of the ALS genes (acetolactate synthase) is the target for both sulfonylurea and imidazolinone herbicides (Lee, et al. (1987), *EMBO*, 7, 1241). Protoplasts resistant to these herbicides have been isolated and shown to contain mutations in one of the two ALS loci. A 10 kb SpeI fragment of the ALS-II gene (Lee, et al. (1988), supra; Mazur, et al. (1987), *Plant Phys.*, 85, 1110) is subcloned into the negative selection vector, pIC-19R/TK. A neo^r gene, engineered for expression in plant cells with regulating sequences from the mannopine synthase gene for the TI plasmid is inserted into the EcoRI site in the coding region of the ALS-II. This PNS vector is transferred to the C3 tobacco cell line (Chalef, et al. (1984), *Science*, 223, 1148), carrying a chlorsulfuron^r allele in Als-I.

They are then selected for Neo^r, GanC^r. Those cells surviving selection are screened by DNA blots for candidates containing insertions in the ALS-II gene.

Having described the preferred embodiments of the present invention, it will appear to those ordinarily skilled in the art that various modifications may be made to the disclosed embodiments, and that such modifications are intended to be within the scope of the present invention.

What is claimed is:

1. A positive-negative selection (PNS) vector for modifying a target DNA sequence contained in the genomes of murine embryonic stem cells, said PNS vector comprising:

a first homologous vector DNA sequence capable of homologous recombination with a first region of said target DNA sequence,

a positive selection marker DNA sequence capable of conferring a positive selection characteristic in said cells,

a second homologous vector DNA sequence capable of homologous recombination with a second region of said target DNA sequence, and

a negative selection marker DNA sequence, capable of conferring a negative selection characteristic in said cells, thereby allowing killing of said cells, but substantially incapable of homologous recombination with said target DNA sequence,

wherein the spatial order of said sequences in said PNS vector is: said first homologous vector DNA sequence, said positive selection marker DNA sequence, said second homologous vector DNA sequence and said negative selection marker DNA sequence as shown in FIG. **1**,

wherein the 5'-3' orientation of said first homologous vector sequence relative to said second homologous vector sequence is the same as the 5'-3' orientation of said first region relative to said second region of said target sequence;

wherein the vector is capable of modifying said target DNA sequence by homologous recombination of said first homologous vector DNA sequence with said first region of said target sequence and of said second homologous vector DNA sequence with said second region of said target sequence.

2. The PNS vector of claim **1** wherein said target DNA

contains exons and introns and said positive selection marker DNA sequence further contains the exon-intron and intron-exon splicing sequences for an intron in said target DNA.

3. The PNS vector of claim 2 wherein said first or said second homologous vector DNA sequence contains at least a portion of an exon wherein one or more nucleotides have been substituted, deleted or inserted.

4. The PNS vector of claim 1 wherein said target DNA sequence contains exons and introns and said first and second homologous vector DNA sequences contain different portions of the same exon of said target DNA sequence.

5. The PNS vector of claim 1 wherein said PNS vector has a length between 20 kb and 50 kb.

6. The PNS vector of claim 1 wherein said first and said second homologous vector DNA sequences have a length between 25 base pairs and 50,000 base pairs each.

7. The PNS vector of claim 1 wherein said first and said second homologous vector DNA sequences have a length between 1,000 base pairs and 15,000 base pairs each.

8. The PNS vector of claim 1 wherein said positive selection marker DNA sequence is selected from the group consisting of DNA sequences encoding neomycin resistance, hygromycin resistance, histidinol resistance, xanthine utilization and bleomycin resistance.

9. The PNS vector of claim 8 wherein said positive selection marker is a DNA sequence encoding neomycin resistance.

10. The PNS vector of claim 1 wherein said negative selection marker DNA sequence is selected from the group consisting of DNA sequences encoding Hprt, gpt, HSV-tk, diphtheria toxin, ricin toxin and cytosine deaminase.

11. The PNS vector of claim 1 wherein said negative selection marker is a DNA sequence encoding HSV-tk.

12. The PNS vector of claim 1 wherein said first or said second homologous vector DNA sequence comprises a DNA sequence having a modification of said target DNA sequence.

13. The PNS vector of claim 12 wherein said modification is an insertion of one or more nucleotides.

14. The PNS vector of claim 1 wherein said first or said second homologous vector DNA sequence encodes the correction of a genetic defect in said target DNA sequence.

15. The PNS vector of claim 14 wherein said genetic defect in said target DNA comprises the insertion of one or more nucleotides in said target DNA sequence.

16. The PNS vector of claim 15 wherein said genetic defect is associated with hemoglobinopathies, deficiencies in circulatory factors, intracellular enzymes or extracellular enzymes.

17. The PNS vector of claim 1 wherein said PNS vector is linear.

18. The PNS vector of claim 1 wherein said PNS vector is closed circular.

19. A method for enriching for a transformed murine embryonic stem cell containing a modification in a target DNA sequence in the genome of said cell comprising:

(a) transfecting cells capable of mediating homologous recombination with a positive-negative selection vector comprising:

a first homologous vector DNA sequence capable of homologous recombination with a first region of said target DNA sequence,

a positive selection marker DNA sequence capable of conferring a positive selection characteristic in said cells,

a second homologous vector DNA sequence capable of homologous recombination with a second region of said target DNA sequence, and

a negative selection marker DNA sequence, capable of conferring a negative selection characteristic in said cells, thereby allowing killing of said cells but substantially incapable of homologous recombination with said target DNA sequence,

wherein the spatial order of said sequences in said PNS vector is: said first homologous vector DNA sequence, said positive selection marker DNA sequence, said second homologous vector DNA sequence and said negative selection marker DNA sequence as shown in FIG. 1,

wherein the 5'-3' orientation of said first homologous vector sequence relative to said second homologous vector sequence is the same as the 5'-3' orientation of said first region relative to said second region of said target sequence;

wherein the vector is capable of modifying the target DNA sequence by homologous recombination of said first and second homologous vector sequences with the first and second regions of said target sequence;

(b) selecting for transformed cells in which said positive-negative selection vector has integrated into said target DNA sequence by homologous recombination by sequentially or simultaneously selecting against transformed cells containing said negative selection marker and selecting for cells containing said positive selection marker; and

(c) analyzing the DNA of transformed cells surviving the selecting step to identify a cell containing the modification.

20. The method of claim 19 wherein said target DNA contains exons and introns and said positive selection marker DNA sequence further contains the exon-intron and intron-exon splice sequences for an intron in said target DNA sequence.

21. The method of claim 19 wherein said first or said second homologous vector DNA sequence contains at least a portion of an exon of said target DNA sequence wherein one or more nucleotides of said target sequence have been substituted, deleted or inserted.

22. The method of claim 19 wherein said target DNA sequence contains exons and introns and said first and second homologous vector DNA sequences contain different portions of the same exon of said target DNA sequence.

23. The method of claim 19 wherein said PNS vector has a length between 20 kb and 50 kb.

24. The method of claim 19 wherein said first and said second homologous vector DNA sequences have a length between 1,000 base pairs and 15,000 base pairs each.

25. The method of claim 19 wherein said positive selection marker DNA sequence is selected from the group consisting of DNA sequences encoding neomycin resistance, hygromycin resistance, histidinol resistance, xanthine utilization and bleomycin resistance.

26. The method of claim 25 wherein said positive selection marker is a DNA sequence encoding neomycin resistance..

27. The method of claim 19 wherein said negative selection marker DNA sequence is selected from the group consisting of DNA sequences encoding Hprt, gpt, HSV-tk, diphtheria toxin, ricin toxin or cytosine deaminase.

28. The method of claim 27 wherein said negative selection marker is a DNA sequence encoding HSV-tk.

29. The method of claim 19 wherein said first or said

second homologous vector DNA sequence in said PNS vector further comprises a DNA sequence having a modification of said target DNA sequence.

30. The method of claim **29** wherein said modification is a substitution, insertion or deletion of one or more nucleotides.

31. The PNS method of claim **19** wherein said first or said second homologous vector DNA sequences in said PNS vector encodes the correction of a genetic defect in said target DNA.

32. The method of claim **31** wherein said genetic defect in said target DNA comprises the insertion of one or more nucleotides in said target DNA sequence.

33. The method of claim **32** wherein said genetic defect is associated with hemoglobinopathies, deficiencies in circulatory factors, extracellular enzymes or intracellular enzymes.

34. The method of claim **19** wherein said PNS vector is linear.

35. The method of claim **19** wherein said PNS vector is closed circular.

36. The PNS vector of claim **1**, wherein said target DNA sequence is a gene.

37. The PNS vector of claim **1**, wherein said target DNA sequence is a regulatory sequence.

38. The PNS vector of claim **12**, wherein said modification is a deletion of one or more nucleotides in said target DNA sequence.

39. The PNS vector of claim **14**, wherein said genetic defect in said target DNA comprises the deletion of one or more nucleotides in said target DNA sequence.

40. The PNS vector of claim **12**, wherein said modification is a substitution of one or more nucleotides in said target DNA sequence.

41. The PNS vector of claim **14**, wherein said genetic defect in said target DNA comprises the substitution of one or more nucleotides in said target DNA sequence.

42. The method of claim **31**, wherein said genetic defect in said target DNA comprises the deletion of one or more nucleotides in said target DNA sequence.

43. The method of claim **31**, wherein said genetic defect in said target DNA comprises the substitution of one or more nucleotides in said target DNA sequence.

44. The method of claim **19** wherein said vector is a sequence replacement vector and said first and second homologous vector DNA sequences comprise contiguous first and second regions in said target DNA sequence.

* * * * *

EXHIBIT 4

US 5888732A

United States Patent [19]

Hartley et al.

[11] Patent Number: 5,888,732

[45] Date of Patent: Mar. 30, 1999

[54] **RECOMBINATIONAL CLONING USING ENGINEERED RECOMBINATION SITES**

[75] Inventors: **James L. Hartley**, Frederick; **Michael A. Brasch**, Gaithersburg, both of Md.

[73] Assignee: **Life Technologies, Inc.**, Rockville, Md.

[21] Appl. No.: **663,002**

[22] Filed: **Jun. 7, 1996**

Related U.S. Application Data

[63] Continuation-in-part of Ser. No. 486,139, Jun. 7, 1995, abandoned.

[51] **Int. Cl.**[6] C12Q 1/68; C12P 19/34; C12N 15/63; C07H 21/04

[52] **U.S. Cl.** 435/6; 435/91.42; 435/320.1; 536/23.1; 536/24.2

[58] **Field of Search** 435/6, 69.1, 91.1, 435/9.4, 172.1, 320.1; 536/23.1, 24.1

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,673,640	6/1987	Backman	435/69.1
5,434,066	7/1995	Bebee et al.	435/172.3
5,591,609	1/1997	Auerbach	435/91.2

FOREIGN PATENT DOCUMENTS

0 300 422 A2	1/1989	European Pat. Off. .
WO 93/19172	9/1993	WIPO .

OTHER PUBLICATIONS

Snaith, M.R., et al., "Multiple cloning sites carrying loxP and FRT recognition sites for the Cre and Flp site-specific recombinases," *Genes 166(1)*:173–174 (Dec. 1, 1995).

Hasan et. al.. *Escherichia coli* geneome targeting, I. Cre–Lox mediated invitro generation of ori– plasmids adn their invivo chromasomal integration and retreival. Gene. vol.. 150:51–56, Dec. 8, 1994.

Abremski, K. et al., "Bacteriophage P1 Cre–loxP Site-specific Recombination: Site-specific DNA Topoisomerase Activity of the Cre Recombination Protein," *J. Biol. Chem. 261(1)*:391–396 (1986).

Adams, D. E. et al., "Cre–lox Recombination in *Escherichia coli* Cells: Mechanistic Differences from the in Vitro Reaction," *J. Mol. Biol. 226*:661–673 (1992).

Araki, H. et al., "Site–specific Recombinase, R, Encoded by Yeast Plasmid pSR1," *J. Mol. Biol. 225*:25–37 (1992).

Argos, P. et al., "The integrase family of site–specific recombinases: regional similarities and global diversity," *EMBO J. 52(2)*:433–440 (1986).

Boyd, A. C., "Turbo cloning: a fast, efficient method for cloning PCR products and other bluntended DNA fragments into plasmids," *Nucl. Acis Res. 21(4)*:817–821 (1993).

Broach, J. R. et al., "Recombination within the Yeast Plasmid 2μ Circle is Site–Specific," *Cell 29*:227–234 (1982).

Campbell, A. M., "Chromosomal Insertion Sites for Phages and Plasmids," *J. Bacteriol. 174(23)*:7495–7499 (1992).

Devine, S. E. and J. D. Boeke, "Efficient integration of artificial transposons into plasmid targets in vitro: a useful tool for DNA mapping, sequencing and genetic analysis," *Nucl. Acids Res 22(18)*:3765–3772 (Sep. 1994).

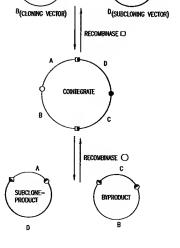
(List continued on next page.)

Primary Examiner—Nancy Degen
Assistant Examiner—William Sandals
Attorney, Agent, or Firm—Sterne, Kessler, Goldstein & Fox PLLC

[57] **ABSTRACT**

Recombinational cloning is provided by the use of nucleic acids, vectors and methods, in vitro and in vivo, for moving or exchanging segments of DNA molecules using engineered recombination sites and recombination proteins to provide chimeric DNA molecules that have the desired characteristic(s) and/or DNA segment(s).

47 Claims, 44 Drawing Sheets



OTHER PUBLICATIONS

Ferguson, J. et al., "Construction and characterization of three yeast–*Escherichia coli* shuttle vectors designed for rapid subcloning of yeast genes on small DNA fragments," *Gene 16*:191–197 (1981).

Filutowicz, M. et al., "Purification of the *Escherichia coli* integration host factor (IHF) in one chromatographic step," *Gene 147*:149–150 (Sep. 1994).

Hasan, N. and W. Szybalski, "Control of cloned gene expression by promoter inversion in vivo: construction of improved vectors with a multiple cloning site and the P_{inc} promoter," *Gene 56*:145–151 (1987).

Hashimoto–Gotoh, T. et al., "Improved vector, pHSG664, for direct streptomycin–resistance selection: cDNA cloning with G:C–tailing procedure and subcloning of double–digested DNA fragments," *Gene 41*:125–128 (1986).

Hoess, R. H. et al., "The role of the loxP spacer region in P1 site–specific recombination," *Nucl. Acids Res. 14(5)*:2287–2300 (1986).

Hoess, R. H. and K. Abremski, "The Cre–lox Recombination System," in: *Nucleic Acids and Molecular Biology, vol. 4*, ed. by Eckstein, F. and D. M. J. Lilley, Springer–Verlag, Berlin, pp. 99–109 (1990).

Landy, A., "Mechanistic and structural complexity in the site–specific recombination pathways of Int and FLP," *Curr. Op. Genet. Develop, 3*:699–707 (1993).

Landy, A., "Dynamic, Structural, and Regulatory Aspects of λ Site–Specific Recombination," *Annu. Rev. Biochem. 58*:913–949 (1989).

Luckow, V. A. et al., "Efficient Generation of Infectious Recombinant Baculoviruses by Site–Specific Transposon–Mediated Insertion of Foreign Genes into a Baculovirus Genome Propagated in *Escherichia coli*," *J. Virol. 67(8)*:4566–4579 (1993).

Maeser, S. and R. Kahmann, "The Gin recombinase of phage Mu can catalyse site–specific recombination in plant protoplasts," *Mol. Gen. Genet.* 230:170–176 (1991).

Mahillon, J. et al., "Is231 and other *Bacillus thuringiensis* transposable elements: a review," *Genetica 93*:13–26 (Nov. 1994).

Nash, H. A. et al., "Role of homology in site–specific recombination of bacteriophage λ: Evidence against joining of cohesive ends," *Proc. Natl. Acad. Sci. USA 84*:4049–4053 (1987).

Nash, H. A., "Purification and Properties of the Bacteriophage Lambda Int Protein," *Meth. Enzymol. 100*:210–216 (1983).

Oberto, J. et al., "A segment of the phage HK022 chromosome is a mosiac of other lambdoid chromosomes," *Nucl. Acids Res. 22(3)*:354–356 (Feb. 1994).

Palazzolo, M. J. et al., "Phage lambda cDNA cloning vectors for substractive hybridization, fusion–protein synthesis and Cre–loxP automatic plasmid subcloning," *Gene 88*:25–36 (1990).

Pósfai, G. et al., "In vivo excision and amplification of large segments of the *Escherichia coli* genome," *Nucl. Acids Res. 22(12)*:2392–2398 (Jun. 1994).

Qian, X. et al., "Reactions between Half– and Full–FLP Recombination Target Sites: A Model System for Analyzing Early Steps in FLP Protein–Mediated Site–Specific Recombination," *J. Biol. Chem 267(11)*:7794–7805(1992).

Sauer, B., "Site–specific recombination: developments and applications," *Curr. Op. Biotechnol. 5*:521–527 (Oct. 1994).

Schlake, T. and J. Bode, "Use of Mutated FLP Recognition Target (FRT) Sites for the Exchange of Expression Cassettes at Defined Chromosomal Loci," *Biochemistry 33*:12746–12751 (Nov. 1994).

Sizemore, C. et al., "Quantitative analysis of Tn10 Tet repressor binding to a complete set of tet operator mutants," *Nucl. Acids Res. 18(10)*:2875–2880 (1990).

Waterhouse, P. et al., "Combinatorial infection and in vivo recombination: a strategy for making large phage antibody repertoires," *Nucl. Acids Res. 21(9)*:2265–2266 (1993).

Weisberg, R. A. and A. Landy, "Site–specific Recombination in Phage Lambda," in: Lambda II, Hendrix, R. W. et al., eds., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y., pp. 211–250 (1983).

Brunelli, J.P. and Pall, M.L., "Lambda/Plasmid Vector Construction by in Vivo cre/lox–Mediated Recombination," *BioTechniques 16*(6):1061–1064 (Jun. 1994).

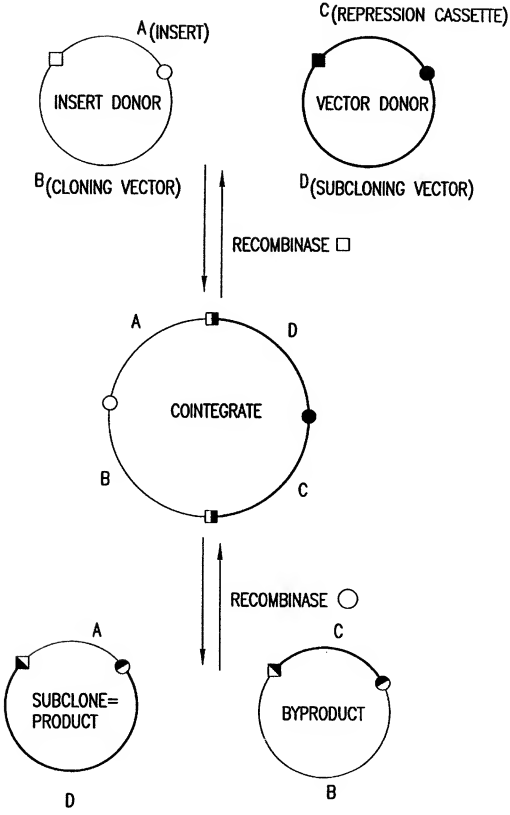

A (INSERT)
INSERT DONOR
C (REPRESSION CASSETTE)
VECTOR DONOR
B (CLONING VECTOR)
D (SUBCLONING VECTOR)
RECOMBINASE □
A
D
COINTEGRATE
B
C
RECOMBINASE ○
A
C
SUBCLONE= PRODUCT
BYPRODUCT
D
B

FIG.1

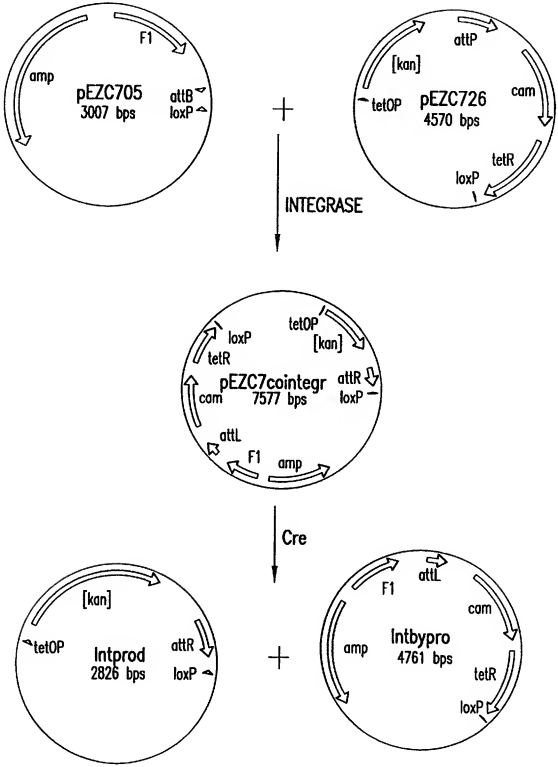

F1
amp
pEZC705
3007 bps
attB
loxP
attP
[kan]
tetOP
pEZC726
4570 bps
cam
tetR
loxP
INTEGRASE
tetOP
[kan]
loxP
tetR
pEZC7cointegr
7577 bps
attR
loxP
cam
attL
F1
amp
Cre
[kan]
tetOP
Intprod
2826 bps
attR
loxP
attL
F1
cam
amp
Intbypro
4761 bps
tetR
loxP

FIG.2A

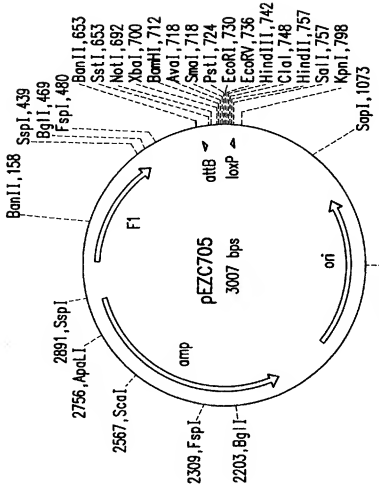

pEZC705
3007 bps
amp
F1
ori
attB
loxP
BanII,158
SspI,439
BglI,469
FspI,480
BanII,653
SstI,653
NotI,692
XbaI,700
BamHI,712
AvaI,718
SmaI,718
PstI,724
EcoRI,730
EcoRV,736
HindIII,742
ClaI,748
HindII,757
SalI,757
KpnI,798
SapI,1073
1510,ApaLI
2203,BglI
2309,FspI
2567,ScaI
2756,ApaLI
2891,SspI

FIG.2B

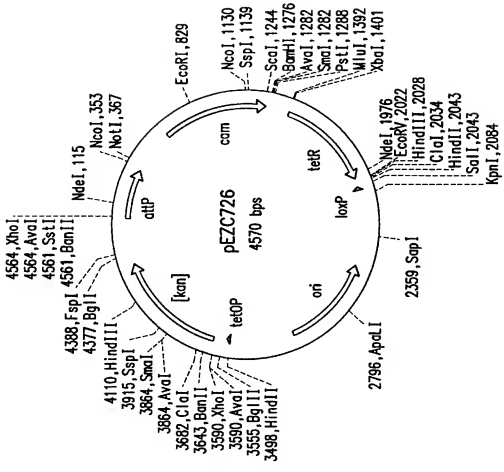

pEZC726
4570 bps
cam
tetR
loxP
ori
tetOP
[kan]
dtIP
EcoRI, 829
NcoI, 1130
SspI, 1139
ScaI, 1244
BamHI, 1276
AvaI, 1282
SmaI, 1282
PstI, 1288
MluI, 1392
XbaI, 1401
NdeI, 1976
EcoRV, 2022
HindIII, 2028
ClaI, 2034
HindII, 2043
SalI, 2043
KpnI, 2084
SapI
2359
ApaLI
2796
HindIII
3498
BglII
3555
AvaI
3590
XhoI
3590
BamII
3643
ClaI
3682
AvaI
3864
SmaI
3864
SspI
3915
HindIII
4110
BglI
4377
FspI
4388
BamII
4561
SstI
4561
AvaI
4564
XhoI
4564
NdeI, 115
NotI, 367
NcoI, 353

FIG.2C

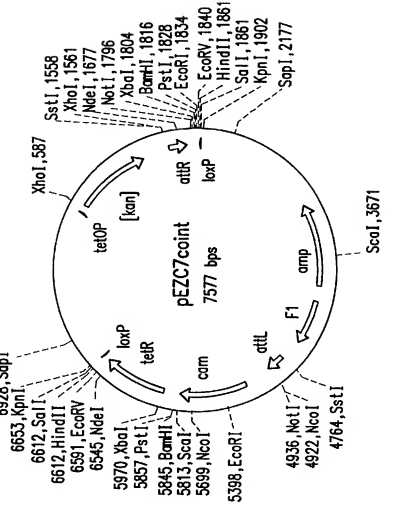

pEZC7coint
7577 bps
tetOP
[kan]
attR
loxP
amp
F1
attL
cam
tetR
loxP
XhoI,587
SstI,1558
XhoI,1561
NdeI,1677
NotI,1796
XbaI,1804
BamHI,1816
PstI,1828
EcoRI,1834
EcoRV,1840
HindII,1861
SalI,1861
KpnI,1902
SapI,2177
ScaI,3671
4764,SstI
4922,NcoI
4936,NotI
5398,EcoRI
5699,NcoI
5813,ScaI
5845,BamHI
5857,PstI
5970,XbaI,
6545,NdeI
6591,EcoRV
6612,HindII
6612,SalI,
6653,KpnI
6928,SapI

FIG.2D

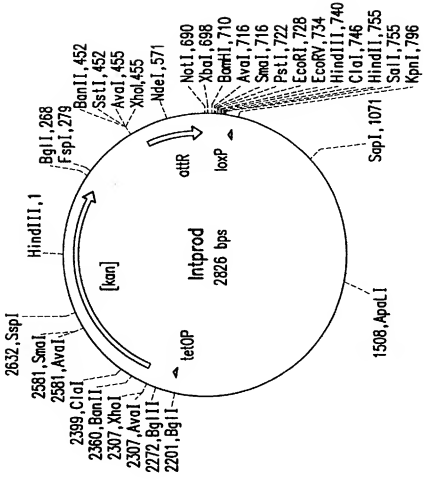

HindIII,1
BglI,268
FspI,279
BanII,452
SstI,452
AvaI,455
XhoI,455
NdeI,571
NotI,690
XbaI,698
BamHI,710
AvaI,716
SmaI,716
PstI,722
EcoRI,728
EcoRV,734
HindIII,740
ClaI,746
HindI,755
SalI,755
KpnI,796
SapI,1071
1508,ApaLI
2632,SspI
2581,SmaI
2581,AvaI
2399,ClaI
2360,BanII
2307,XhoI
2307,AvaI
2272,BglII
2201,BglI
[kan]
Intprod
2826 bps
attR
loxP
tetOP

FIG.2E

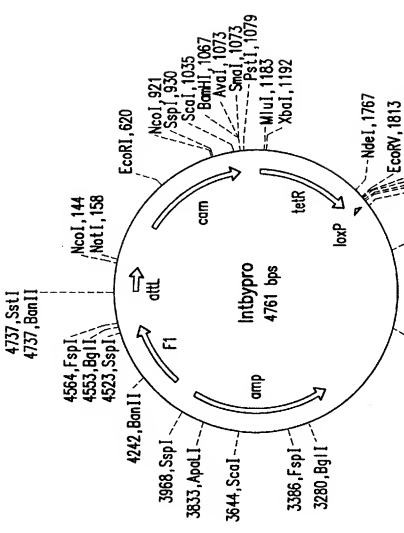

Intbypro
4761 bps
cam
tetR
loxP
attL
F1
amp
EcoRI, 620
NcoI, 921
SspI, 930
ScaI, 1035
BamHI, 1067
AvaI, 1073
SmaI, 1073
PstI, 1079
MluI, 1183
XbaI, 1192
NdeI, 1767
EcoRV, 1813
HindIII, 1819
ClaI, 1825
HindII, 1834
SalI, 1834
KpnI, 1875
SapI, 2150
2587, ApaLI
3386, FspI
3280, BglI
3644, ScaI
3833, ApaLI
3968, SspI
4242, BanII
4564, FspI
4553, BglI
4523, SspI
4737, SstI
4737, BamII
NcoI, 144
NotI, 158

FIG.2F

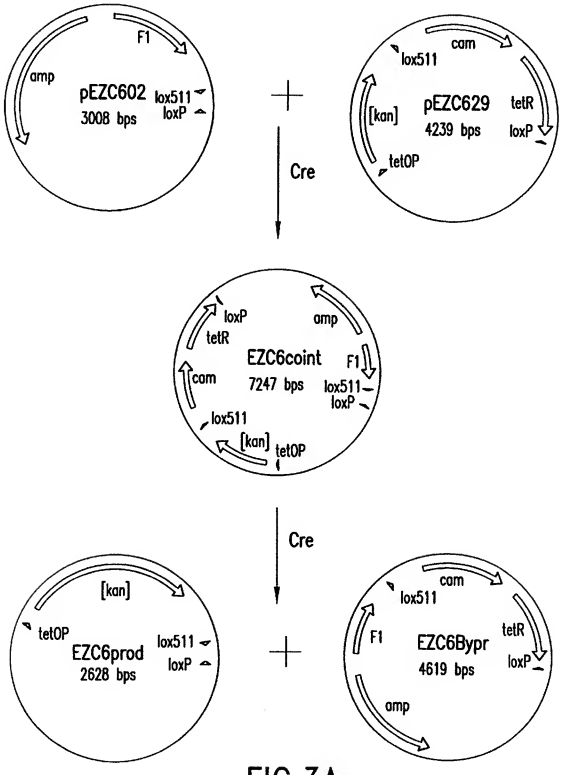

F1
amp
pEZC602
3008 bps
lox511
loxP
cam
lox511
[kan]
pEZC629
4239 bps
tetR
loxP
tetOP
Cre
loxP
tetR
amp
EZC6coint
7247 bps
F1
cam
lox511
loxP
lox511
[kan]
tetOP
Cre
[kan]
tetOP
EZC6prod
2628 bps
lox511
loxP
cam
lox511
F1
EZC6Bypr
4619 bps
tetR
loxP
amp

FIG.3A

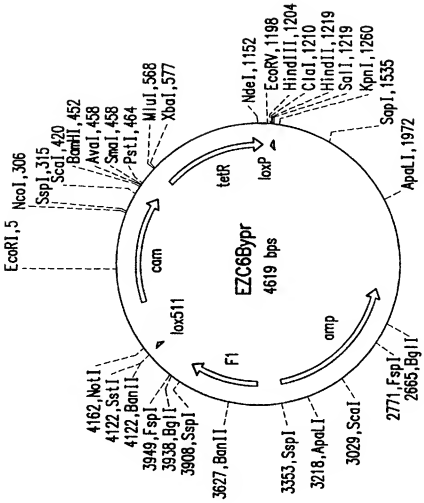

EcoRI,5
NcoI,306
SspI,315
ScaI,420
BamHI,452
AvaI,458
SmaI,458
PstI,464
MluI,568
XbaI,577
NdeI,1152
EcoRV,1198
HindIII,1204
ClaI,1210
HindII,1219
SalI,1219
KpnI,1260
SapI,1535
ApaLI,1972
2665,BglI
2771,FspI
3029,ScaI
3218,ApaLI
3353,SspI
3627,BanII
3908,SspI
3938,BglI
3949,FspI
4122,BanII
4122,SstI
4162,NotI
lox511
cam
tetR
loxP
EZC6Bypr
4619 bps
F1
amp

FIG.3B

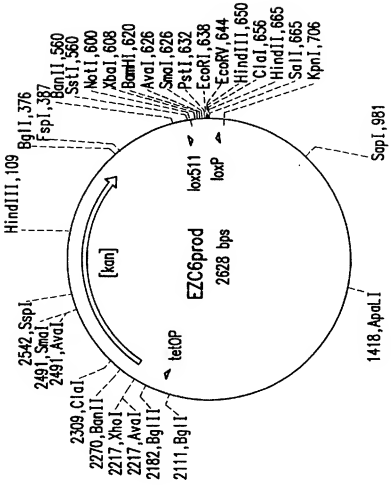

HindIII,109
BglII,376
FspI,387
BanII,560
SstI,560
NotI,600
XbaI,608
BamHI,620
AvaI,626
SmaI,626
PstI,632
EcoRI,638
EcoRV,644
HindIII,650
ClaI,656
HindII,665
SalI,665
KpnI,706
SspI,981
1418,ApaLI
2542,SspI
2491,SmaI
2491',AvaI
2309,ClaI
2270,BanII
2217,XhoI
2217,AvaI
2182,BglII'
2111,BglI'
tetOP
[kan]
EZC6prod
2628 bps
lox511
loxP

FIG.3C

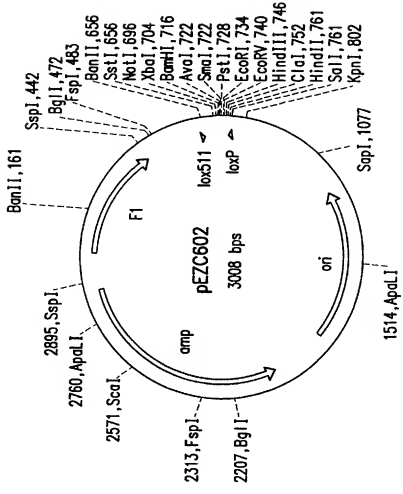

BanII,656
SstI,656
NotI,696
XbaI,704
BamHI,716
AvaI,722
SmaI,722
PstI,728
EcoRI,734
EcoRV,740
HindIII,746
ClaI,752
HindII,761
SalI,761
KpnI,802
SspI,442
BglI,472
FspI,483
BanII,161
SapI,1077
1514,ApaLI
2895,SspI
2760,ApaLI
2571,ScaI
2313,FspI
2207,BglI
F1
lox511
loxP
pEZC602
3008 bps
amp
ori

FIG.3D

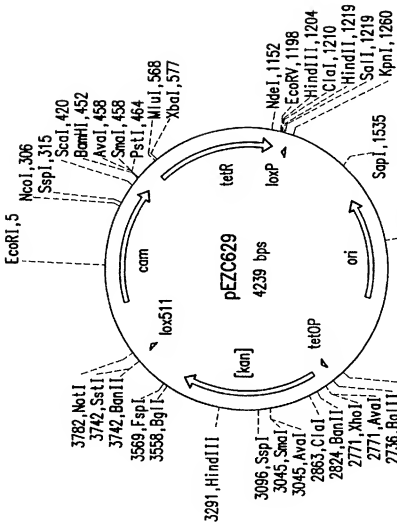

pEZC629
4239 bps
cam
tetR
ori
[kan]
tetOP
lox511
loxP
EcoRI,5
NcoI,306
SspI,315
ScaI,420
BamHI,452
AvaI,458
SmaI,458
PstI,464
MluI,568
XbaI,577
NdeI,1152
EcoRV,1198
HindIII,1204
ClaI,1210
HindII,1219
SalI,1219
KpnI,1260
SapI,1535
ApaLI,1972
3782,NotI
3742,SstII
3742,BanII
3569,FspI
3558,BglI
3291,HindIII
3096,SspI
3045,SmaI
3045,AvaI
2863,ClaI
2824,BanII
2771,XhoI
2771,AvaII
2736,BglII
2665,BglI

FIG.3E

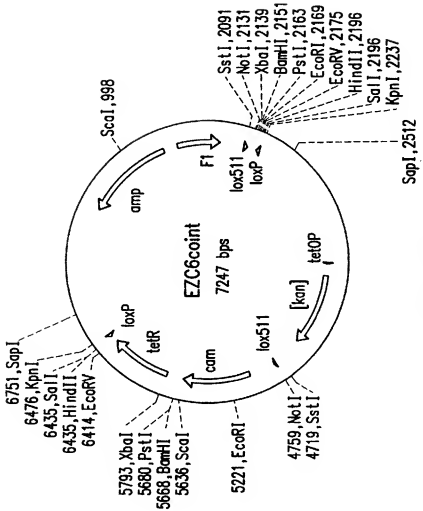

EZC6coint
7247 bps
amp
ScaI,998
F1
lox511
loxP
SstI,2091
NotI,2131
XbaI,2139
BamHI,2151
PstI,2163
EcoRI,2169
EcoRV,2175
HindII,2196
SalI,2196
KpnI,2237
SapI,2512
tetOP
[kan]
lox511
cam
tetR
loxP
6751,SapI
6476,KpnI
6435,SalI
6435,HindII
6414,EcoRV
5793,XbaI
5680,PstI
5668,BamHI
5636,ScaI
5221,EcoRI
4759,NotI
4719,SstI

FIG.3F

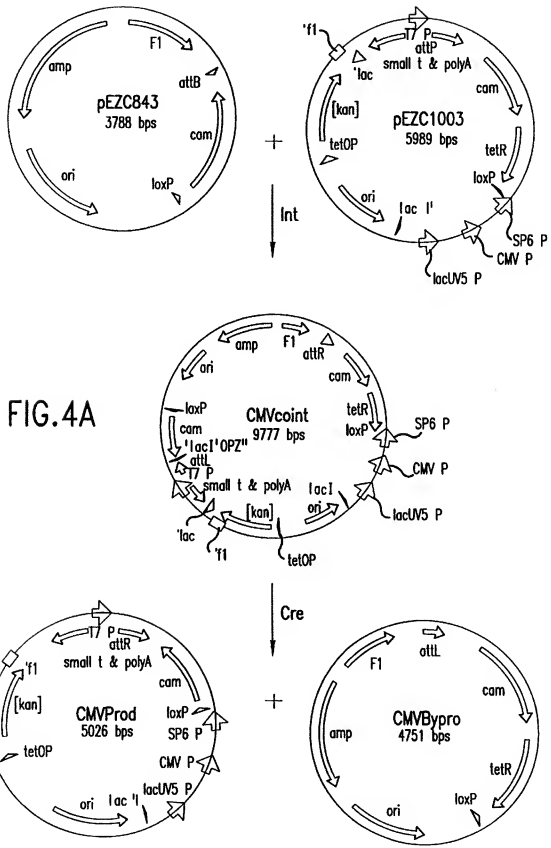

pEZC843
3788 bps
F1
amp
attB
cam
ori
loxP
'f1
T7 P
attP
'lac
small t & polyA
cam
[kan]
pEZC1003
5989 bps
tetOP
tetR
loxP
ori
lac I'
SP6 P
CMV P
lacUV5 P
Int
CMVcoint
9777 bps
amp
F1
attR
ori
cam
loxP
cam
'lacI'OPZ"
attL
T7 P
small t & polyA
[kan]
'lac
'f1
tetOP
tetR
loxP
SP6 P
CMV P
lacI
ori
lacUV5 P
Cre
CMVProd
5026 bps
T7 P
attR
small t & polyA
'f1
[kan]
cam
loxP
SP6 P
tetOP
CMV P
ori
lac I'
lacUV5 P
CMVBypro
4751 bps
attL
F1
cam
amp
tetR
loxP
ori

FIG.4A

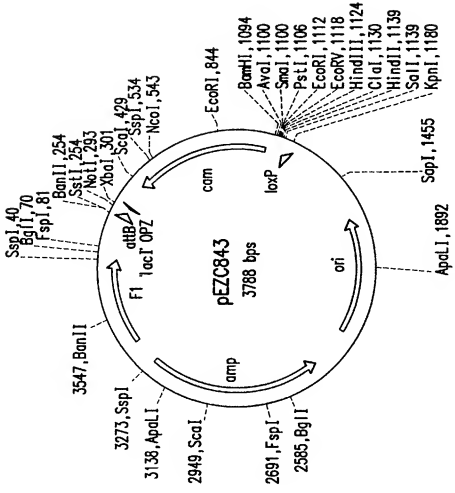

pEZC843
3788 bps
amp
ori
cam
loxP
F1
attB
1acI' OPZ
EcoRI,844
BamHI,1094
AvaI,1100
SmaI,1100
PstI,1106
EcoRI,1112
EcoRV,1118
HindIII,1124
ClaI,1130
HindII,1139
SalI,1139
KpnI,1180
SgpI,1455
ApaLI,1892
NcoI,343
Sspl,534
ScaI,429
XbaI,301
NotI,293
SstI,254
BanII,254
FspI,81
BglI,70
SspI,40
3547,BanII
3273,SspI
3138,ApaLI
2949,ScaI
2691,FspI
2585,BglI

FIG.4B

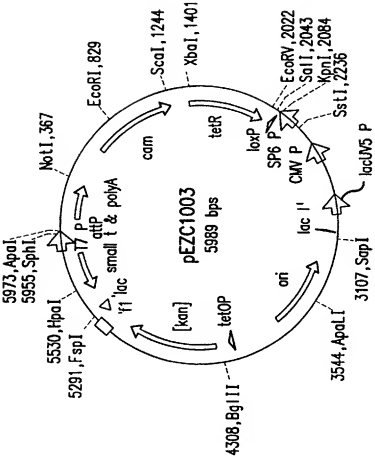

pEZC1003
5989 bps
cam
tetR
ori
[kan]
tetOP
'lac
small t & polyA
dHP
f1
lac I''
loxP
SP6 P
CMV P
lacUV5 P
NotI,367
EcoRI,829
ScaI,1244
XbaI,1401
EcoRV,2022
SalI,2043
KpnI,2084
SstI,2236
5973,ApaI
5955,SphI
5530,HpaI
5291,FspI
4308,BglII
3544,ApaLI
3107,SapI

FIG.4C

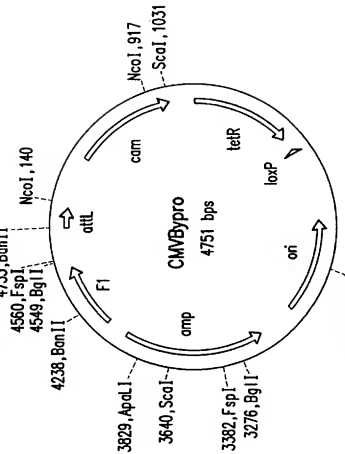

NcoI,917
ScaI,1031
cam
tetR
loxP
ori
2583,ApaLI
3382,FspI
3276,BglI
3640,ScaI
3829,ApaLI
4238,BamII
4560,FspI
4549,BglI
4733,BamII
NcoI,140
attL
F1
amp
CMVBypro
4751 bps

FIG.4D

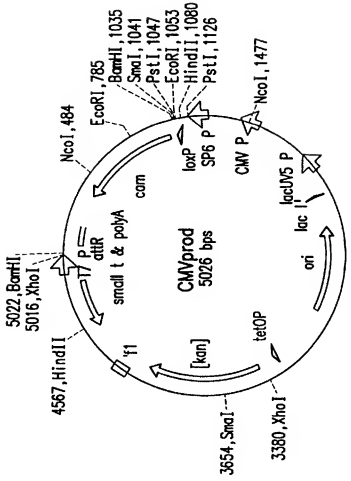

EcoRI, 785
BamHI, 1035
SmaI, 1041
PstI, 1047
EcoRI, 1053
HindII, 1080
PstI, 1126
NcoI, 1477
NcoI, 484
5022, BamHI,
5016, XhoI
4567, HindII
3654, SmaI
3380, XhoI
CMVprod
5026 bps
loxP
SP6 P
CMV P
lacUV5 P
lac
ori
tetOP
[kan]
f1
small t & polyA
dtiR
P
T7
cam

FIG.4E

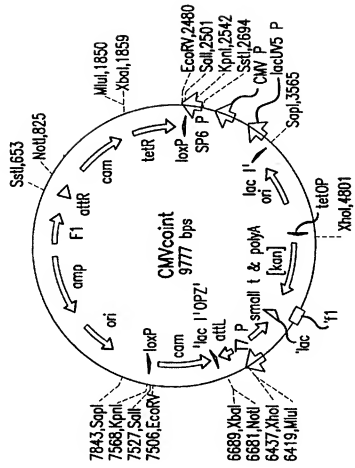

CMVcoint
9777 bps
loxP
cam
'lac I'OPZ'
dttL
P
small t & polyA
[kan]
'lac
'f1
ori
lac I'
tetOP
Xhol,4801
lacUV5 P
SopI,3365
CMV P
SstII,2694
KpnI,2542
SalI,2501
EcoRV,2480
SP6 P
loxP
tetR
cam
dttR
F1
amp
ori
MluI,1850
XbaI,1859
SstII,653
NotI,825
7843,SopI
7568,KpnI
7527,SalI
7506,EcoRV
6689,XbaI
6681,NotI
6437,XhoI
6419,MluI

FIG.4F

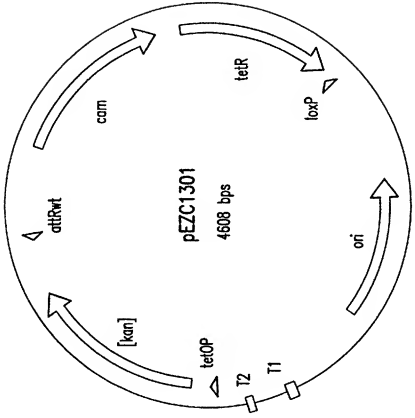


FIG.5A

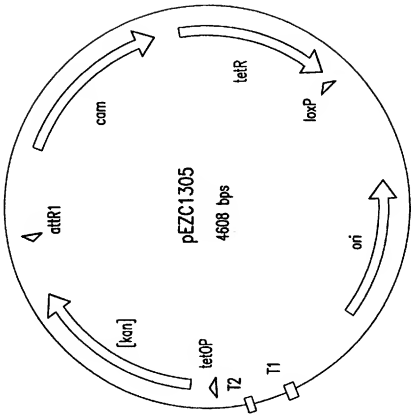


FIG.5B

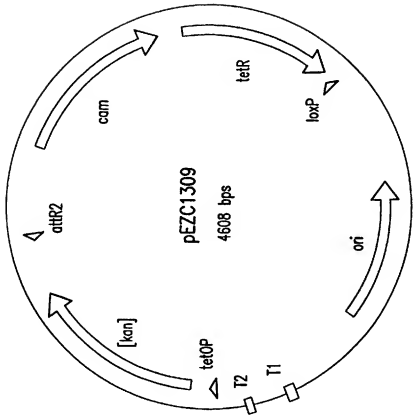


FIG.5C

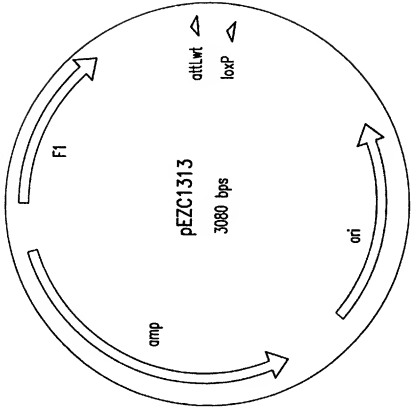


FIG.5D

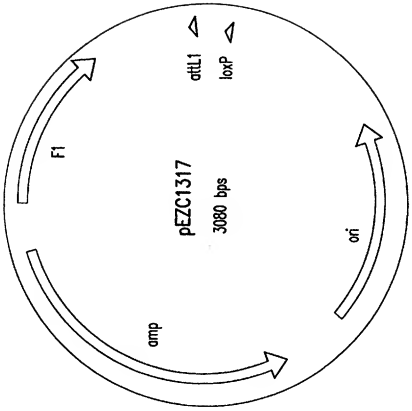


FIG.5E

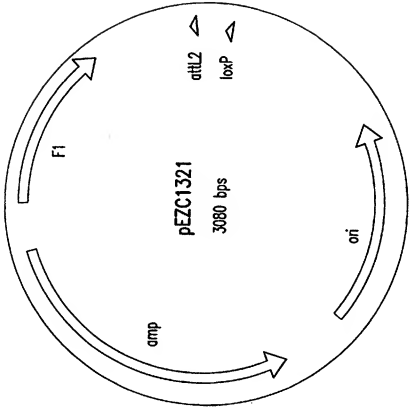


FIG.5F

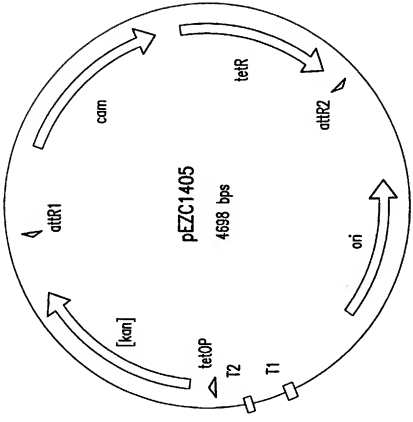


FIG.5G

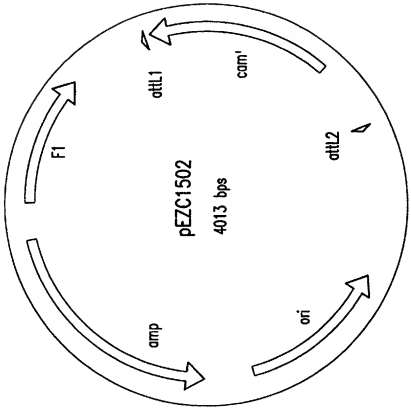


FIG.5H

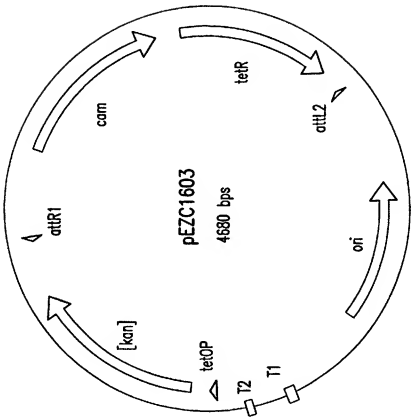


FIG. 6A

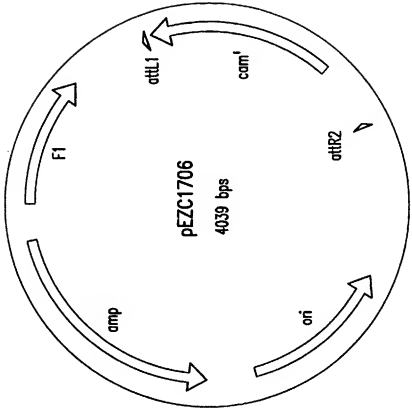


FIG.6B

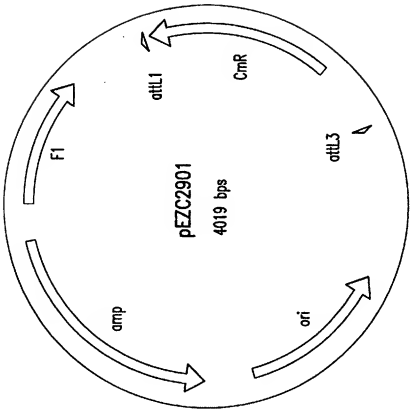


FIG.7A

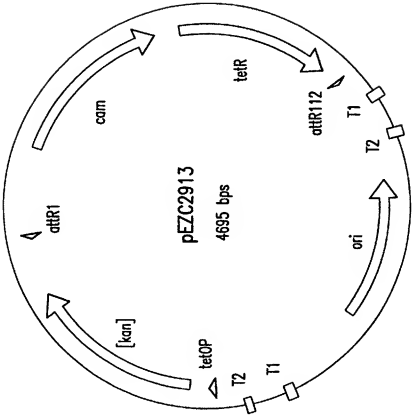

pEZC2913
4695 bps
attR1
cam
tetR
attR112
T1
T2
ori
T1
T2
tetOP
[kan]

FIG.7B

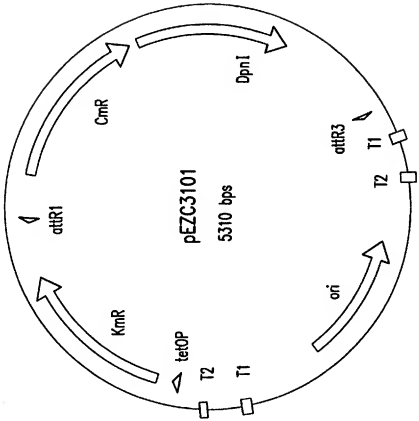

pEZC3101
5310 bps
attR1
CmR
Dpn1
attR3
T1
T2
ori
T2
T1
tetOP
KmR

FIG.7C

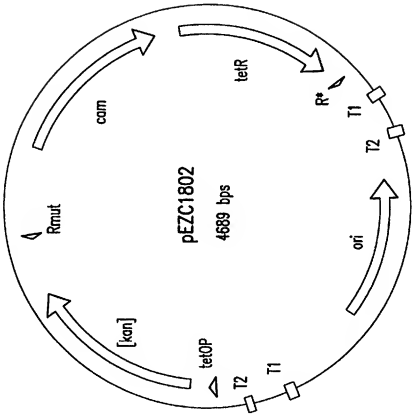


FIG.7D

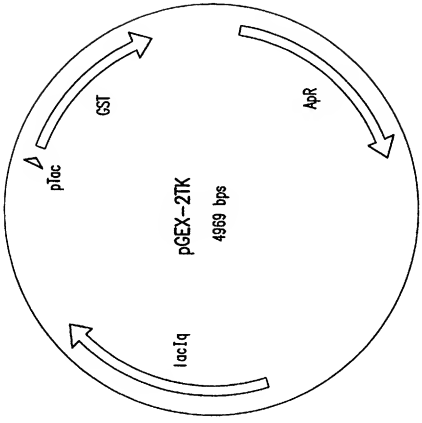


FIG.8A

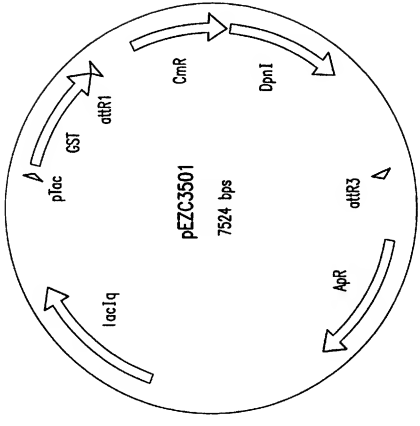


FIG.8B

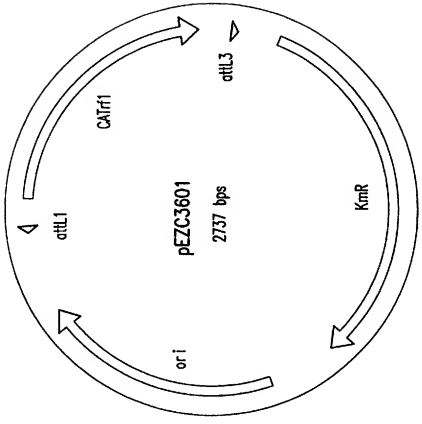


FIG.8C

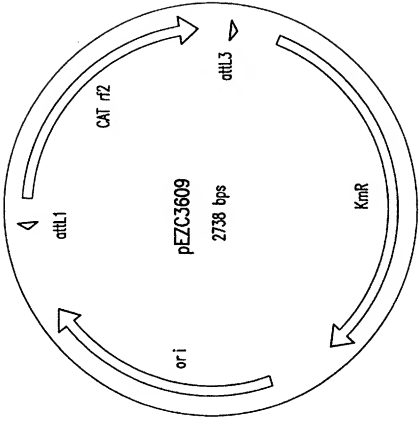


FIG.8D

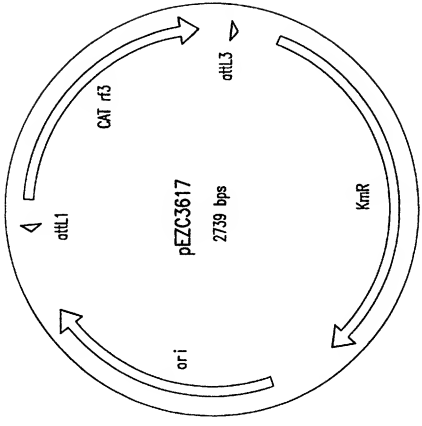


FIG.8E

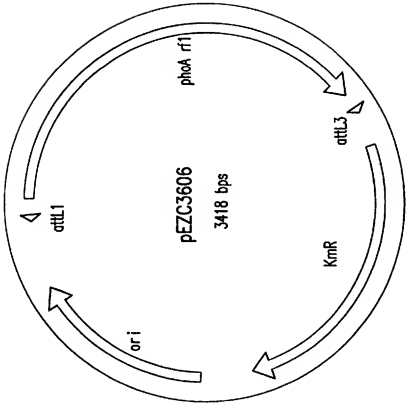


FIG.8F

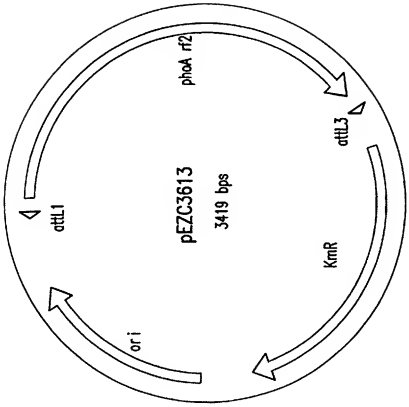


FIG.8G

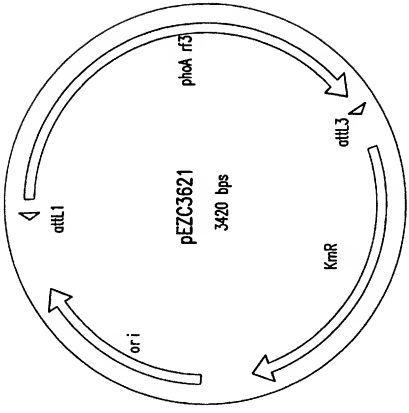


FIG.8H

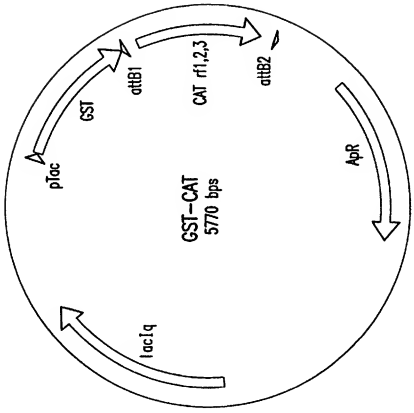


FIG.8I

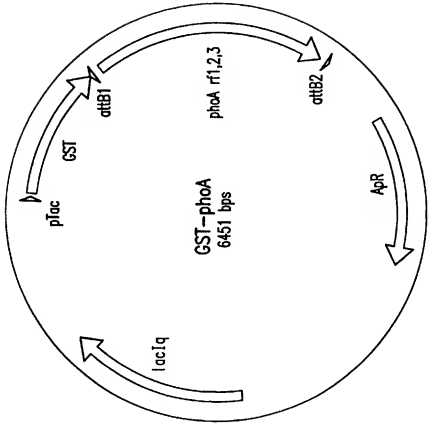


FIG.8J

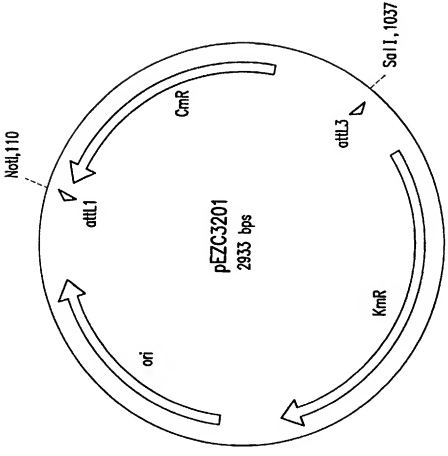


FIG. 8K

RECOMBINATIONAL CLONING USING ENGINEERED RECOMBINATION SITES

CROSS-REFERENCE TO RELATED APPLICATIONS

The present application is a continuation-in-part of U.S. application Ser. No. 08/486,139, filed Jun. 7, 1995, now abandoned which application is entirely incorporated herein by reference.

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to recombinant DNA technology. DNA and vectors having engineered recombination sites are provided for use in a recombinational cloning method that enables efficient and specific recombination of DNA segments using recombination proteins. The DNAs, vectors and methods are useful for a variety of DNA exchanges, such as subcloning of DNA, in vitro or in vivo.

2. Related Art

Site specific recombinases. Site specific recombinases are enzymes that are present in some viruses and bacteria and have been characterized to have both endonuclease and ligase properties. These recombinases (along with associated proteins in some cases) recognize specific sequences of bases in DNA and exchange the DNA segments flanking those segments. The recombinases and associated proteins are collectively referred to as "recombination proteins" (see, e.g., Landy, A., *Current Opinion in Biotechnology* 3:699–707 (1993)).

Numerous recombination systems from various organisms have been described. See, e.g., Hoess et al., *Nucleic Acids Research* 14(6):2287 (1986); Abremski et al., *J. Biol. Chem.* 261(1):391 (1986); Campbell, *J. Bacteriol.* 174(23) :7495 (1992); Qian et al., *J. Biol. Chem.* 267(1 1):7794 (1992); Araki et al., *J. Mol. Biol.* 225(1):25 (1992); Maeser and Kahnmann (1991) *Mol. Gen. Genet.* 230:170–176).

Many of these belong to the integrase family of recombinases (Argos et al. *EMBO J.* 5:433–440 (1986)). Perhaps the best studied of these are the Integrase/att system from bacteriophage λ (Landy, A. *Current Opinions in Genetics and Devel.* 3:699–707 (1993)), the Cre/loxP system from bacteriophage P1 (Hoess and Abremski (1990) In *Nucleic Acids and Molecular Biology*, vol. 4. Eds.: Eckstein and Lilley, Berlin-Heidelberg: Springer-Verlag; pp. 90–109), and the FLP/FRT system from the Saccharomyces cerevisiae 2 μ circle plasmid (Broach et al. *Cell* 29:227–234 (1982)).

While these recombination systems have been characterized for particular organisms, the related art has only taught using recombinant DNA flanked by recombination sites, for in viva recombination.

Backman (U.S. Pat. No. 4,673,640) discloses the in vivo use of λ recombinase to recombine a protein producing DNA segment by enzymatic site-specific recombination using wild-type recombination sites attB and attP.

Hasan and Szybalski (*Gene* 56:145–151 (1987)) discloses the use of λ Int recombinase in vivo for intramolecular recombination between wild type attP and attB sites which flank a promoter. Because the orientations of these sites are inverted relative to each other, this causes an irreversible flipping of the promoter region relative to the gene of interest.

Palazzolo et al. *Gene* 88:25–36 (1990), discloses phage lambda vectors having bacteriophage λ arms that contain restriction sites positioned outside a cloned DNA sequence and between wild-type loxP sites. Infection of *E. coli* cells that express the Cre recombinase with these phage vectors results in recombination between the loxP sites and the in vivo excision of the plasmid replicon, including the cloned cDNA.

Pósfai et al. (*Nucl. Acids Res.* 22:2392–2398 (1994)) discloses a method for inserting into genomic DNA partial expression vectors having a selectable marker, flanked by two wild-type FRT recognition sequences. FLP site-specific recombinase as present in the cells is used to integrate the vectors into the genome at predetermined sites. Under conditions where the replicon is functional, this cloned genomic DNA can be amplified.

Bebee et al. (U.S. Pat. No. 5,434,066) discloses the use of site-specific recombinases such as Cre for DNA containing two loxP sites is used for in vivo recombination between the sites.

Boyd (*Nucl. Acids Res.* 21:817–821 (1993)) discloses a method to facilitate the cloning of blunt-ended DNA using conditions that encourage intermolecular ligation to a dephosphorylated vector that contains a wild-type loxP site acted upon by a Cre site-specific recombinase present in *E. coli* host cells.

Waterhouse et al. (PCT No. 93/19172 and *Nucleic Acids Res.* 21 (9):2265 (1993)) disclose an in vivo method where light and heavy chains of a particular antibody were cloned in different phage vectors between loxP and loxP 511 sites and used to transfect new *E. coli* cells. Cre, acting in the host cells on the two parental molecules (one plasmid, one phage), produced four products in equilibrium: two different cointegrates (produced by recombination at either loxP or loxP 511 sites), and two daughter molecules, one of which was the desired product.

In contrast to the other related art, Schlake & Bode (*Biochemistry* 33:12746–12751 (1994)) discloses an in vivo method to exchange expression cassettes at defined chromosomal locations, each flanked by a wild type and a spacer-mutated FRT recombination site. A double-reciprocal crossover was mediated in cultured mammalian cells by using this FLP/FRT system for site-specific recombination.

Transposases. The family of enzymes, the transposases, has also been used to transfer genetic information between replicons. Transposons are structurally variable, being described as simple or compound, but typically encode the recombinase gene flanked by DNA sequences organized in inverted orientations. Integration of transposons can be random or highly specific. Representatives such as Tn7, which are highly site-specific, have been applied to the in vivo movement of DNA segments between replicons (Lucklow et al., *J. Virol.* 67:4566–4579 (1993)).

Devine and Boeke *Nucl. Acids Res.* 22:3765–3772 (1994), discloses the construction of artificial transposons for the insertion of DNA segments, in vitro, into recipient DNA molecules. The system makes use of the integrase of yeast TY1 virus-like particles. The DNA segment of interest is cloned, using standard methods, between the ends of the transposon-like element TY1. In the presence of the TY1 integrase, the resulting element integrates randomly into a second target DNA molecule.

DNA cloning. The cloning of DNA segments currently occurs as a daily routine in many research labs and as a prerequisite step in many genetic analyses. The purpose of these clonings is various, however, two general purposes can be considered: (1) the initial cloning of DNA from large DNA or RNA segments (chromosomes, YACs, PCR fragments, mRNA, etc.), done in a relative handful of known

vectors such as pUC, pGem, pBlueScript, and (2) the subcloning of these DNA segments into specialized vectors for functional analysis. A great deal of time and effort is expended both in the initial cloning of DNA segments and in the transfer of DNA segments from the initial cloning vectors to the more specialized vectors. This transfer is called subcloning.

The basic methods for cloning have been known for many years and have changed little during that time. A typical cloning protocol is as follows:

(1) digest the DNA of interest with one or two restriction enzymes;

(2) gel purify the DNA segment of interest when known;

(3) prepare the vector by cutting with appropriate restriction enzymes, treating with alkaline phosphatase, gel purify etc., as appropriate;

(4) ligate the DNA segment to vector, with appropriate controls to estimate background of uncut and self-ligated vector;

(5) introduce the resulting vector into an *E. coli* host cell;

(6) pick selected colonies and grow small cultures overnight;

(7) make DNA minipreps; and

(8) analyze the isolated plasmid on agarose gels (often after diagnostic restriction enzyme digestions) or by PCR.

The specialized vectors used for subcloning DNA segments are functionally diverse. These include but are not limited to: vectors for expressing genes in various organisms; for regulating gene expression; for providing tags to aid in protein purification or to allow tracking of proteins in cells; for modifying the cloned DNA segment (e.g., generating deletions); for the synthesis of probes (e.g, riboprobes); for the preparation of templates for DNA sequencing; for the identification of protein coding regions; for the fusion of various protein-coding regions; to provide large amounts of the DNA of interest, etc. It is common that a particular investigation will involve subcloning the DNA segment of interest into several different specialized vectors.

As known in the art, simple subclonings can be done in one day (e.g., the DNA segment is not large and the restriction sites are compatible with those of the subcloning vector). However, many other subclonings can take several weeks, especially those involving unknown sequences, long fragments, toxic genes, unsuitable placement of restriction sites, high backgrounds, impure enzymes, etc. Subcloning DNA fragments is thus often viewed as a chore to be done as few times as possible.

Several methods for facilitating the cloning of DNA segments have been described, e.g., as in the following references.

Ferguson, J., et al. *Gene* 16:191 (1981), discloses a family of vectors for subcloning fragments of yeast DNA. The vectors encode kanamycin resistance. Clones of longer yeast DNA segments can be partially digested and ligated into the subcloning vectors. If the original cloning vector conveys resistance to ampicillin, no purification is necessary prior to transformation, since the selection will be for kanamycin.

Hashimoto-Gotoh, T., et al. *Gene* 41:125 (1986), discloses a subcloning vector with unique cloning sites within a streptomycin sensitivity gene; in a streptomycin-resistant host, only plasmids with inserts or deletions in the dominant sensitivity gene will survive streptomycin selection.

Accordingly, traditional subcloning methods, using restriction enzymes and ligase, are time consuming and relatively unreliable. Considerable labor is expended, and if two or more days later the desired subclone can not be found among the candidate plasmids, the entire process must then be repeated with alternative conditions attempted. Although site specific recombinases have been used to recombine DNA in vivo, the successful use of such enzymes in vitro was expected to suffer from several problems. For example, the site specificities and efficiencies were expected to differ in vitro; topologically-linked products were expected; and the topology of the DNA substrates and recombination proteins was expected to differ significantly in vitro (see, e.g., Adams et al, *J. Mol. Biol.* 226:661–73 (1992)). Reactions that could go on for many hours in vivo were expected to occur in significantly less time in vitro before the enzymes became inactive. Multiple DNA recombination products were expected in the biological host used, resulting in unsatisfactory reliability, specificity or efficiency of subcloning. In vitro recombination reactions were not expected to be sufficiently efficient to yield the desired levels of product.

Accordingly, there is a long felt need to provide an alternative subcloning system that provides advantages over the known use of restriction enzymes and ligases.

SUMMARY OF THE INVENTION

The present invention provides nucleic acid, vectors and methods for obtaining chimeric nucleic acid using recombination proteins and engineered recombination sites, in vitro or in vivo. These methods are highly specific, rapid, and less labor intensive than what is disclosed or suggested in the related background art. The improved specificity, speed and yields of the present invention facilitates DNA or RNA subcloning, regulation or exchange useful for any related purpose. Such purposes include in vitro recombination of DNA segments and in vitro or in vivo insertion or modification of transcribed, replicated, isolated or genomic DNA or RNA.

The present invention relates to nucleic acids, vectors and methods for moving or exchanging segments of DNA using at least one engineered recombination site and at least one recombination protein to provide chimeric DNA molecules which have the desired characteristic(s) and/or DNA segment(s). Generally, one or more parent DNA molecules are recombined to give one or more daughter molecules, at least one of which is the desired Product DNA segment or vector. The invention thus relates to DNA, RNA, vectors and methods to effect the exchange and/or to select for one or more desired products.

One embodiment of the present invention relates to a method of making chimeric DNA, which comprises

(a) combining in vitro or in vivo

(i) an Insert Donor DNA molecule, comprising a desired DNA segment flanked by a first recombination site and a second recombination site, wherein the first and second recombination sites do not recombine with each other;

(ii) a Vector Donor DNA molecule containing a third recombination site and a fourth recombination site, wherein the third and fourth recombination sites do not recombine with each other; and

(iii) one or more site specific recombination proteins capable of recombining the first and third recombinational sites and/or the second and fourth recombinational sites;

thereby allowing recombination to occur, so as to produce at least one Cointegrate DNA molecule, at least one desired Product DNA molecule which comprises said desired DNA segment, and optionally a Byproduct DNA molecule; and then, optionally,

(b) selecting for the Product or Byproduct DNA molecule.

Another embodiment of the present invention relates to a kit comprising a carrier or receptacle being compartmentalized to receive and hold therein at least one container, wherein a first container contains a DNA molecule comprising a vector having at least two recombination sites flanking a cloning site or a Selectable marker, as described herein. The kit optionally further comprises:

(i) a second container containing a Vector Donor plasmid comprising a subcloning vector and/or a Selectable marker of which one or both are flanked by one or more engineered recombination sites; and/or

(ii) a third container containing at least one recombination protein which recognizes and is capable of recombining at least one of said recombination sites.

Other embodiments include DNA and vectors useful in the methods of the present invention. In particular, Vector Donor molecules are provided in one embodiment, wherein DNA segments within the Vector Donor are separated either by, (i) in a circular Vector Donor, at least two recombination sites, or (ii) in a linear Vector Donor, at least one recombination site, where the recombination sites are preferably engineered to enhance specificity or efficiency of recombination.

One Vector Donor embodiment comprises a first DNA segment and a second DNA segment, the first or second segment comprising a Selectable marker. A second Vector Donor embodiment comprises a first DNA segment and a second DNA segment, the first or second DNA segment comprising a toxic gene. A third Vector Donor embodiment comprises a first DNA segment and a second DNA segment, the first or second DNA segment comprising an inactive fragment of at least one Selectable marker, wherein the inactive fragment of the Selectable marker is capable of reconstituting a functional Selectable marker when recombined across the first or second recombination site with another inactive fragment of at least one Selectable marker.

The present recombinational cloning method possesses several advantages over previous in vivo methods. Since single molecules of recombination products can be introduced into a biological host, propagation of the desired Product DNA in the absence of other DNA molecules (e.g., starting molecules, intermediates, and by-products) is more readily realized. Reaction conditions can be freely adjusted in vitro to optimize enzyme activities. DNA molecules can be incompatible with the desired biological host (e.g, YACs, genomic DNA, etc.), can be used. Recombination proteins from diverse sources can be employed, together or sequentially.

Other embodiments will be evident to those of ordinary skill in the art from the teachings contained herein in combination with what is known to the art.

BRIEF DESCRIPTION OF THE FIGURES

FIG. 1 depicts one general method of the present invention, wherein the starting (parent) DNA molecules can be circular or linear. The goal is to exchange the new subcloning vector D for the original cloning vector B. It is desirable in one embodiment to select for AD and against all the other molecules, including the Cointegrate. The square and circle are sites of recombination: e.g., loxP sites, att sites, etc. For example, segment D can contain expression signals, new drug markers, new origins of replication, or specialized functions for mapping or sequencing DNA.

FIG. 2A depicts an in vitro method of recombining an Insert Donor plasmid (here, pEZC705) with a Vector Donor plasmid (here, pEZC726), and obtaining Product DNA and Byproduct daughter molecules. The two recombination sites are attP and loxP on the Vector Donor. On one segment defined by these sites is a kanamycin resistance gene whose promoter has been replaced by the tetOP operator/promoter from transposon Tn10. See Sizemore et al., *Nucl. Acids Res.* 18(10):2875 (1990). In the absence of tet repressor protein, *E. coli* RNA polymerase transcribes the kanamycin resistance gene from the tetOP. If tet repressor is present, it binds to tetOP and blocks transcription of the kanamycin resistance gene. The other segment of pEZC726 has the tet repressor gene expressed by a constitutive promoter. Thus cells transformed by pEZC726 are resistant to chloramphenicol, because of the chloramphenicol acetyl transferase gene on the same segment as tetR, but are sensitive to kanamycin. The recombinase-mediated reactions result in separation of the tetR gene from the regulated kanamycin resistance gene. This separation results in kanamycin resistance in cells receiving only the desired recombination products. The first recombination reaction is driven by the addition of the recombinase called Integrase. The second recombination reaction is driven by adding the recombinase Cre to the Cointegrate (here, pEZC7 Cointegrate).

FIG. 2B depicts a restriction map of pEZC705.

FIG. 2C depicts a restriction map of pEZC726.

FIG. 2D depicts a restriction map of pEZC7 Cointegrate.

FIG. 2E depicts a restriction map of Intprod.

FIG. 2F depicts a restriction map of Intbypro.

FIG. 3A depicts an in vitro method of recombining an Insert Donor plasmid (here, pEZC602) with a Vector Donor plasmid (here, pEZC629), and obtaining Product (here, EZC6prod) and Byproduct (here, EZC6Bypr) daughter molecules. The two recombination sites are loxP and loxP 511. One segment of pEZC629 defined by these sites is a kanamycin resistance gene whose promoter has been replaced by the tetOP operator/promoter from transposon Tn10. In the absence of tet repressor protein, *E. coli* RNA polymerase transcribes the kanamycin resistance gene from the tetOP. If tet repressor is present, it binds to tetOP and blocks transcription of the kanamycin resistance gene. The other segment of pEZC629 has the tet repressor gene expressed by a constitutive promoter. Thus cells transformed by pEZC629 are resistant to chloramphenicol, because of the chloramphenicol acetyl transferase gene on the same segment as tetR, but are sensitive to kanamycin. The reactions result in separation of the tetR gene from the regulated kanamycin resistance gene. This separation results in kanamycin resistance in cells receiving the desired recombination product. The first and the second recombination events are driven by the addition of the same recombinase, Cre.

FIG. 3B depicts a restriction map of EZC6Bypr.

FIG. 3C depicts a restriction map of EZC6prod.

FIG. 3D depicts a restriction map of pEZC602.

FIG. 3E depicts a restriction map of pEZC629.

FIG. 3F depicts a restriction map of EZC6coint.

FIG. 4A depicts an application of the in vitro method of recombinational cloning to subclone the chloramphenicol acetyl transferase gene into a vector for expression in eukaryotic cells. The Insert Donor plasmid, pEZC843, is comprised of the chloramphenicol acetyl transferase gene of *E. coli*, cloned between loxP and attB sites such that the loxP site is positioned at the 5'-end of the gene. The Vector Donor plasmid, pEZC1003, contains the cytomegalovirus eukaryotic promoter apposed to a loxP site. The supercoiled

plasmids were combined with lambda Integrase and Cre recombinase in vitro. After incubation, competent *E. coli* cells were transformed with the recombinational reaction solution. Aliquots of transformations were spread on agar plates containing kanamycin to select for the Product molecule (here CMVProd).

FIG. 4B depicts a restriction map of pEZC843.

FIG. 4C depicts a restriction map of pEZC1003.

FIG. 4D depicts a restriction map of CMVBypro.

FIG. 4E depicts a restriction map of CMVProd.

FIG. 4F depicts a restriction map of CMVcoint.

FIG. 5A depicts a vector diagram of pEZC1301.

FIG. 5B depicts a vector diagram of pEZC1305.

FIG. 5C depicts a vector diagram of pEZC1309.

FIG. 5D depicts a vector diagram of pEZC1313.

FIG. 5E depicts a vector diagram of pEZC1317.

FIG. 5F depicts a vector diagram of pEZC1321.

FIG. 5G depicts a vector diagram of pEZC1405.

FIG. 5H depicts a vector diagram of pEZC1502.

FIG. 6A depicts a vector diagram of pEZC1603.

FIG. 6B depicts a vector diagram of pEZC1706.

FIG. 7A depicts a vector diagram of pEZC2901.

FIG. 7B depicts a vector diagram of pEZC2913

FIG. 7C depicts a vector diagram of pEZC3101.

FIG. 7D depicts a vector diagram of pEZC1802.

FIG. 8A depicts a vector diagram of pGEX-2TK.

FIG. 8B depicts a vector diagram of pEZC3501.

FIG. 8C depicts a vector diagram of pEZC3601.

FIG. 8D depicts a vector diagram of pEZC3609.

FIG. 8E depicts a vector diagram of pEZC3617.

FIG. 8F depicts a vector diagram of pEZC3606.

FIG. 8G depicts a vector diagram of pEZC3613.

FIG. 8H depicts a vector diagram of pEZC3621.

FIG. 8I depicts a vector diagram of GST-CAT.

FIG. 8J depicts a vector diagram of GST-phoA.

FIG. 8K depicts a vector diagram of pEZC3201.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

It is unexpectedly discovered in the present invention that subcloning reactions can be provided using recombinational cloning. Recombination cloning according to the present invention uses DNAs, vectors and methods, in vitro and in vivo, for moving or exchanging segments of DNA molecules using engineered recombination sites and recombination proteins. These methods provide chimeric DNA molecules that have the desired characteristic(s) and/or DNA segment(s).

The present invention thus provides nucleic acid, vectors and methods for obtaining chimeric nucleic acid using recombination proteins and engineered recombination sites, in vitro or in vivo. These methods are highly specific, rapid, and less labor intensive than what is disclosed or suggested in the related background art. The improved specificity, speed and yields of the present invention facilitates DNA or RNA subcloning, regulation or exchange useful for any related purpose. Such purposes include in vitro recombination of DNA segments and in vitro or in vivo insertion or modification of transcribed, replicated, isolated or genomic DNA or RNA.

Definitions

In the description that follows, a number of terms used in recombinant DNA technology are utilized extensively. In order to provide a clear and consistent understanding of the specification and claims, including the scope to be given such terms, the following definitions are provided.

Byproduct: is a daughter molecule (a new clone produced after the second recombination event during the recombinational cloning process) lacking the DNA which is desired to be subcloned.

Cointegrate: is at least one recombination intermediate DNA molecule of the present invention that contains both parental (starting) DNA molecules. It will usually be circular. In some embodiments it can be linear.

Host: is any prokaryotic or eukaryotic organism that can be a recipient of the recombinational cloning Product. A "host," as the term is used herein, includes prokaryotic or eukaryotic organisms that can be genetically engineered. For examples of such hosts, see Maniatis et al., *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1982).

Insert: is the desired DNA segment (segment A of FIG. 1) which one wishes to manipulate by the method of the present invention. The insert can have one or more genes.

Insert Donor: is one of the two parental DNA molecules of the present invention which carries the Insert. The Insert Donor DNA molecule comprises the Insert flanked on both sides with recombination signals. The Insert Donor can be linear or circular. In one embodiment of the invention, the Insert Donor is a circular DNA molecule and further comprises a cloning vector sequence outside of the recombination signals (see FIG. 1).

Product: is one or both the desired daughter molecules comprising the A and D or B and C sequences which are produced after the second recombination event during the recombinational cloning process (see FIG. 1). The Product contains the DNA which was to be cloned or subcloned.

Promoter: is a DNA sequence generally described as the 5'-region of a gene, located proximal to the start codon. The transcription of an adjacent DNA segment is initiated at the promoter region. A repressible promoter's rate of transcription decreases in response to a repressing agent. An inducible promoter's rate of transcription increases in response to an inducing agent. A constitutive promoter's rate of transcription is not specifically regulated, though it can vary under the influence of general metabolic conditions.

Recognition sequence: Recognition sequences are particular DNA sequences which a protein, DNA, or RNA molecule (e.g., restriction endonuclease, a modification methylase, or a recombinase) recognizes and binds. For example, the recognition sequence for Cre recombinase is loxP which is a 34 base pair sequence comprised of two 13 base pair inverted repeats (serving as the recombinase binding sites) flanking an 8 base pair core sequence. See FIG. 1 of Sauer, B., *Current Opinion in Biotechnology* 5:521–527 (1994). Other examples of recognition sequences are the attB, attP, attL, and attR sequences which are recognized by the recombinase enzyme λ Integrase. attB is an approximately 25 base pair sequence containing two 9 base pair core-type Int binding sites and a 7 base pair overlap region. attP is an approximately 240 base pair sequence containing core-type Int binding sites and arm-type Int binding sites as well as sites for auxiliary proteins IHF, FIS, and Xis. See Landy, *Current Opinion in Biotechnology* 3:699–707 (1993). Such sites are also engineered according to the present invention to enhance methods and products.

Recombinase: is an enzyme which catalyzes the exchange of DNA segments at specific recombination sites.

Recombinational Cloning: is a method described herein, whereby segments of DNA molecules are exchanged, inserted, replaced, substituted or modified, in vitro or in vivo.

Recombination proteins: include excisive or integrative proteins, enzymes, co-factors or associated proteins that are involved in recombination reactions involving one or more recombination sites. See, Landy (1994), infra.

Repression cassette: is a DNA segment that contains a repressor of a Selectable marker present in the subcloning vector.

Selectable marker: is a DNA segment that allows one to select for or against a molecule or a cell that contains it, often under particular conditions. These markers can encode an activity, such as, but not limited to, production of RNA, peptide, or protein, or can provide a binding site for RNA, peptides, proteins, inorganic and organic compounds or compositions and the like. Examples of Selectable markers include but are not limited to: (1) DNA segments that encode products which provide resistance against otherwise toxic compounds (e.g., antibiotics); (2) DNA segments that encode products which are otherwise lacking in the recipient cell (e.g., tRNA genes, auxotrophic markers); (3) DNA segments that encode products which suppress the activity of a gene product; (4) DNA segments that encode products which can be readily identified (e.g., phenotypic markers such as β-galactosidase, green fluorescent protein (GFP), and cell surface proteins); (5) DNA segments that bind products which are otherwise detrimental to cell survival and/or function; (6) DNA segments that otherwise inhibit the activity of any of the DNA segments described in Nos. 1–5 above (e.g., antisense oligonucleotides); (7) DNA segments that bind products that modify a substrate (e.g. restriction endonucleases); (8) DNA segments that can be used to isolate a desired molecule (e.g specific protein binding sites); (9) DNA segments that encode a specific nucleotide sequence which can be otherwise non-functional (e.g., for PCR amplification of subpopulations of molecules); and/or (10) DNA segments, which when absent, directly or indirectly confer sensitivity to particular compounds.

Selection scheme: is any method which allows selection, enrichment, or identification of a desired Product or Product (s) from a mixture containing the Insert Donor, Vector Donor, and/or any intermediates, (e.g. a Cointegrate) Byproducts. The selection schemes of one preferred embodiment have at least two components that are either linked or unlinked during recombinational cloning. One component is a Selectable marker. The other component controls the expression in vitro or in vivo of the Selectable marker, or survival of the cell harboring the plasmid carrying the Selectable marker. Generally, this controlling element will be a repressor or inducer of the Selectable marker, but other means for controlling expression of the Selectable marker can be used. Whether a repressor or activator is used will depend on whether the marker is for a positive or negative selection, and the exact arrangement of the various DNA segments, as will be readily apparent to those skilled in the art. A preferred requirement is that the selection scheme results in selection of or enrichment for only one or more desired Products. As defined herein, to select for a DNA molecule includes (a) selecting or enriching for the presence of the desired DNA molecule, and (b) selecting or enriching against the presence of DNA molecules that are not the desired DNA molecule.

In one embodiment, the selection schemes (which can be carried out reversed) will take one of three forms, which will be discussed in terms of FIG. 1. The first, exemplified herein with a Selectable marker and a repressor therefor, selects for molecules having segment D and lacking segment C. The second selects against molecules having segment C and for molecules having segment D. Possible embodiments of the second form would have a DNA segment carrying a gene toxic to cells into which the in vitro reaction products are to be introduced. A toxic gene can be a DNA that is expressed as a toxic gene product (a toxic protein or RNA), or can be toxic in and of itself. (In the latter case, the toxic gene is understood to carry its classical definition of "heritable trait".)

Examples of such toxic gene products are well known in the art, and include, but are not limited to, restriction endonucleases (e.g., DpnI) and genes that kill hosts in the absence of a suppressing function, e.g., kicB. A toxic gene can alternatively be selectable in vitro, e.g., a restriction site.

In the second form, segment D carries a Selectable marker. The toxic gene would eliminate transformants harboring the Vector Donor, Cointegrate, and Byproduct molecules, while the Selectable marker can be used to select for cells containing the Product and against cells harboring only the Insert Donor.

The third form selects for cells that have both segments A and D in cis on the same molecule, but not for cells that have both segments in trans on different molecules. This could be embodied by a Selectable marker that is split into two inactive fragments, one each on segments A and D.

The fragments are so arranged relative to the recombination sites that when the segments are brought together by the recombination event, they reconstitute a functional Selectable marker. For example, the recombinational event can link a promoter with a structural gene, can link two fragments of a structural gene, or can link genes that encode a heterodimeric gene product needed for survival, or can link portions of a replicon.

Site-specific recombinase: is a type of recombinase which typically has at least the following four activities: (1) recognition of one or two specific DNA sequences; (2) cleavage of said DNA sequence or sequences; (3) DNA topoisomerase activity involved in strand exchange; and (4) DNA ligase activity to reseal the cleaved strands of DNA. See Sauer, B., *Current Opinions in Biotechnology* 5:521–527 (1994). Conservative site-specific recombination is distinguished from homologous recombination and transposition by a high degree of specificity for both partners. The strand exchange mechanism involves the cleavage and rejoining of specific DNA sequences in the absence of DNA synthesis (Landy, A. (1989) *Ann. Rev. Biochem.* 58:913–949).

Subcloning vector: is a cloning vector comprising a circular or linear DNA molecule which includes an appropriate replicon. In the present invention, the subcloning vector (segment D in FIG. 1) can also contain functional and/or regulatory elements that are desired to be incorporated into the final product to act upon or with the cloned DNA Insert (segment A in FIG. 1). The subcloning vector can also contain a Selectable marker (contained in segment C in FIG. 1).

Vector: is a DNA that provides a useful biological or biochemical property to an Insert. Examples include plasmids, phages, and other DNA sequences which are able to replicate or be replicated in vitro or in a host cell, or to convey a desired DNA segment to a desired location within a host cell. A Vector can have one or more restriction endonuclease recognition sites at which the DNA sequences can be cut in a determinable fashion without loss of an essential biological function of the vector, and into which a DNA fragment can be spliced in order to bring about its replication and cloning. Vectors can further provide primer sites, e.g., for PCR, transcriptional and/or translational ini-

tiation and/or regulation sites, recombinational signals, replicons, Selectable markers, etc. Clearly, methods of inserting a desired DNA fragment which do not require the use of homologous recombination or restriction enzymes (such as, but not limited to, UDG cloning of PCR fragments (U.S. Pat. No. 5,334,575, entirely incorporated herein by reference), T:A cloning, and the like) can also be applied to clone a fragment of DNA into a cloning vector to be used according to the present invention. The cloning vector can further contain a Selectable marker suitable for use in the identification of cells transformed with the cloning vector.

Vector Donor: is one of the two parental DNA molecules of the present invention which carries the DNA segments encoding the DNA vector which is to become part of the desired Product. The Vector Donor comprises a subcloning vector D (or it can be called the cloning vector if the Insert Donor does not already contain a cloning vector) and a segment C flanked by recombination sites (see FIG. 1). Segments C and/or D can contain elements that contribute to selection for the desired Product daughter molecule, as described above for selection schemes. The recombination signals can be the same or different, and can be acted upon by the same or different recombinases. In addition, the Vector Donor can be linear or circular.

Description

One general scheme for an in vitro or in vivo method of the invention is shown in FIG. 1, where the Insert Donor and the Vector Donor can be either circular or linear DNA, but is shown as circular. Vector D is exchanged for the original cloning vector A. It is desirable to select for the daughter vector containing elements A and D and against other molecules, including one or more Cointegrate(s). The square and circle are different sets of recombination sites (e.g., lox sites or att sites). Segment A or D can contain at least one Selection Marker, expression signals, origins of replication, or specialized functions for detecting, selecting, expressing, mapping or sequencing DNA, where D is used in this example.

Examples of desired DNA segments that can be part of Element A or D include, but are not limited to, PCR products, large DNA segments, genomic clones or fragments, cDNA clones, functional elements, etc., and genes or partial genes, which encode useful nucleic acids or proteins. Moreover, the recombinational cloning of the present invention can be used to make ex vivo and in vivo gene transfer vehicles for protein expression and/or gene therapy.

In FIG. 1, the scheme provides the desired Product as containing vectors D and A, as follows. The Insert Donor (containing A and B) is first recombined at the square recombination sites by recombination proteins, with the Vector Donor (containing C and D), to form a Co-integrate having each of A-D-C-B. Next, recombination occurs at the circle recombination sites to form Product DNA (A and D) and Byproduct DNA (C and B). However, if desired, two or more different Co-integrates can be formed to generate two or more Products.

In one embodiment of the present in vitro or in vivo recombinational cloning method, a method for selecting at least one desired Product DNA is provided. This can be understood by consideration of the map of plasmid pEZC726 depicted in FIG. 2. The two exemplary recombination sites are attP and loxP. On one segment defined by these sites is a kanamycin resistance gene whose promoter has been replaced by the tetOP operator/promoter from transposon Tn10. In the absence of tet repressor protein, *E. coli* RNA polymerase transcribes the kanamycin resistance gene from the tetOP. If tet repressor is present, it binds to tetOP and blocks transcription of the kanamycin resistance gene. The other segment of pEZC726 has the tet repressor gene expressed by a constitutive promoter. Thus cells transformed by pEZC726 are resistant to chloramphenicol, because of the chloramphenicol acetyl transferase gene on the same segment as tetR, but are sensitive to kanamycin. The recombination reactions result in separation of the tetR gene from the regulated kanamycin resistance gene. This separation results in kanamycin resistance in cells receiving the desired recombination Product.

Two different sets of plasmids were constructed to demonstrate the in vitro method. One set, for use with Cre recombinase only (cloning vector 602 and subcloning vector 629 (FIG. 3)) contained loxP and loxP 511 sites. A second set, for use with Cre and integrase (cloning vector 705 and subcloning vector 726 (FIG. 2)) contained loxP and att sites. The efficiency of production of the desired daughter plasmid was about 60 fold higher using both enzymes than using Cre alone. Nineteen of twenty four colonies from the Cre-only reaction contained the desired product, while thirty eight of thirty eight colonies from the integrase plus Cre reaction contained the desired product plasmid.

Other Selection Schemes A variety of selection schemes can be used that are known in the art as they can suit a particular purpose for which the recombinational cloning is carried out. Depending upon individual preferences and needs, a number of different types of selection schemes can be used in the recombinational cloning method of the present invention. The skilled artisan can take advantage of the availability of the many DNA segments or methods for making them and the different methods of selection that are routinely used in the art. Such DNA segments include but are not limited to those which encodes an activity such as, but not limited to, production of RNA, peptide, or protein, or providing a binding site for such RNA, peptide, or protein. Examples of DNA molecules used in devising a selection scheme are given above, under the definition of "selection scheme"

Additional examples include but are not limited to:

(i) Generation of new primer sites for PCR (e.g., juxtaposition of two DNA sequences that were not previously juxtaposed);

(ii) Inclusion of a DNA sequence acted upon by a restriction endonuclease or other DNA modifying enzyme, chemical, ribozyme, etc.;

(iii) Inclusion of a DNA sequence recognized by a DNA binding protein, RNA, DNA, chemical, etc.) (e.g., for use as an affinity tag for selecting for or excluding from a population) (Davis, *Nucl. Acids Res.* 24.:702–706 (1996); *J. Virol.* 69. 8027–8034 (1995));

(iv) In vitro selection of RNA ligands for the ribosomal L22 protein associated with Epstein-Barr virus-expressed RNA by using randomized and cDNA-derived RNA libraries;

(vi) The positioning of functional elements whose activity requires a specific orientation or juxtaposition (e.g, (a) a recombination site which reacts poorly in trans, but when placed in cis, in the presence of the appropriate proteins, results in recombination that destroys certain populations of molecules; (e.g., reconstitution of a promoter sequence that allows in vitro RNA synthesis). The RNA can be used directly, or can be reverse transcribed to obtain the desired DNA construct;

(vii) Selection of the desired product by size (e.g., fractionation) or other physical property of the molecule(s); and

(viii) Inclusion of a DNA sequence required for a specific modification (e.g., methylation) that allows its identification.

After formation of the Product and Byproduct in the method of the present invention, the selection step can be carried out either in vitro or in vivo depending upon the particular selection scheme which has been optionally devised in the particular recombinational cloning procedure.

For example, an in vitro method of selection can be devised for the Insert Donor and Vector Donor DNA molecules. Such scheme can involve engineering a rare restriction site in the starting circular vectors in such a way that after the recombination events the rare cutting sites end up in the Byproduct. Hence, when the restriction enzyme which binds and cuts at the rare restriction site is added to the reaction mixture in vitro, all of the DNA molecules carrying the rare cutting site, i.e., the starting DNA molecules, the Cointegrate, and the Byproduct, will be cut and rendered nonreplicable in the intended host cell. For example, cutting sites in segments B and C (see FIG. 1) can be used to select against all molecules except the Product. Alternatively, only a cutting site in C is needed if one is able to select for segment D, e.g, by a drug resistance gene not found on B.

Similarly, an in vitro selection method can be devised when dealing with linear DNA molecules. DNA sequences complementary to a PCR primer sequence can be so engineered that they are transferred, through the recombinational cloning method, only to the Product molecule. After the reactions are completed, the appropriate primers are added to the reaction solution and the sample is subjected to PCR. Hence, all or part of the Product molecule is amplified.

Other in vivo selection schemes can be used with a variety of *E. coli* cell lines. One is to put a repressor gene on one segment of the subcloning plasmid, and a drug marker controlled by that repressor on the other segment of the same plasmid. Another is to put a killer gene on segment C of the subcloning plasmid (FIG. 1). Of course a way must exist for growing such a plasmid, i.e., there must exist circumstances under which the killer gene will not kill. There are a number of these genes known which require particular strains of *E. coli*. One such scheme is to use the restriction enzyme DpnI, which will not cleave unless its recognition sequence GATC is methylated. Many popular common *E. coli* strains methylate GATC sequences, but there are mutants in which cloned DpnI can be expressed without harm.

Of course analogous selection schemes can be devised for other host organisms. For example, the tet repressor/operator of Tn10 has been adapted to control gene expression in eukaryotes (Gossen, M., and Bujard, H., *Proc. Natl. Acad. Sci. USA* 89:5547–5551 (1992)). Thus the same control of drug resistance by the tet repressor exemplified herein can be applied to select for Product in eukaryotic cells.

Recombination Proteins

In the present invention, the exchange of DNA segments is achieved by the use of recombination proteins, including recombinases and associated co-factors and proteins. Various recombination proteins are described in the art. Examples of such recombinases include:

Cre: A protein from bacteriophage P1 (Abremski and Hoess, *J. Biol. Chem.*259 (3):1509–1514 (1984)) catalyzes the exchange (i.e., causes recombination) between 34 bp DNA sequences called loxP (locus of crossover) sites (See Hoess et al, *Nucl. Acids Res.* 14(5):2287 (1986)). Cre is available commercially (Novagen, Catalog No. 69247-1). Recombination mediated by Cre is freely reversible. From thermodynamic considerations it is not surprising that Cre-mediated integration (recombination between two molecules to form one molecule) is much less efficient than Cre-mediated excision (recombination between two loxP sites in the same molecule to form two daughter molecules). Cre works in simple buffers with either magnesium or spermidine as a cofactor, as is well known in the art. The DNA substrates can be either linear or supercoiled. A number of mutant loxP sites have been described (Hoess et al., supra). One of these, loxP 511, recombines with another loxP 511 site, but will not recombine with a loxP site.

Integrase: A protein from bacteriophage lambda that mediates the integration of the lambda genome into the *E. coli* chromosome. The bacteriophage λ Int recombinational proteins promote irreversible recombination between its substrate att sites as part of the formation or induction of a lysogenic state. Reversibility of the recombination reactions results from two independent pathways for integrative and excisive recombination. Each pathway uses a unique, but overlapping, set of the 15 protein binding sites that comprise att site DNAs. Cooperative and competitive interactions involving four proteins (Int, Xis, IHF and FIS) determine the direction of recombination.

Integrative recombination involves the Int and IHF proteins and sites attP (240 bp) and attB (25 bp). Recombination results in the formation of two new sites: attL and attR. Excisive recombination requires Int, IHF, and Xis, and sites attL and attR to generate attP and aftB. Under certain conditions, FIS stimulates excisive recombination. In addition to these normal reactions, it should be appreciated that attP and attB, when placed on the same molecule, can promote excisive recombination to generate two excision products, one with attL and one with attR. Similarly, intermolecular recombination between molecules containing attL and attR, in the presence of Int, IHF and Xis, can result in integrative recombination and the generation attP and attB. Hence, by flanking DNA segments with appropriate combinations of engineered att sites, in the presence of the appropriate recombination proteins, one can direct excisive or integrative recombination, as reverse reactions of each other.

Each of the att sites contains a 15 bp core sequence; individual sequence elements of functional significance lie within, outside, and across the boundaries of this common core (Landy, A., *Ann. Rev. Biochem.* 58:913 (1989)). Efficient recombination between the various att sites requires that the sequence of the central common region be identical between the recombining partners, however, the exact sequence is now found to be modifiable. Consequently, derivatives of the att site with changes within the core are now discovered to recombine as least as efficiently as the native core sequences.

Integrase acts to recombine the attP site on bacteriophage lambda (about 240 bp) with the attB site on the *E. coli* genome (about 25 bp) (Weisberg, R. A. and Landy, A. in *Lambda II*, p. 211 (1983), Cold Spring Harbor Laboratory)), to produce the integrated lambda genome flanked by attL (about 100 bp) and attR (about 160 bp) sites. In the absence of Xis (see below), this reaction is essentially irreversible. The integration reaction mediated by integrase and IHF works in vitro, with simple buffer containing spermidine. Integrase can be obtained as described by Nash, H. A., *Methods of Enzymology* 100:210–216 (1983). IHF can be obtained as described by Filutowicz, M., et al., *Gene* 147:149–150 (1994).

In the presence of the λ protein Xis (excise) integrase catalyzes the reaction of attR and attL to form attP and attB, i.e., it promotes the reverse of the reaction described above. This reaction can also be applied in the present invention.

Other Recombination Systems. Numerous recombination systems from various organisms can also be used, based on the teaching and guidance provided herein. See, e.g., Hoess et al., *Nucleic Acids Research* 14(6):2287 (1986); Abremski et al., *J. Biol. Chem.* 261(1):391 (1986); Campbell, *J. Bacteriol.* 174(23):7495 (1992); Qian et al., *J. Biol. Chem.* 267(11):7794 (1992); Araki et al., *J. Mol. Biol.* 225(1):25 (1992)). Many of these belong to the integrase family of recombinases (Argos et al. *EMBO J.* 5:433–440 (1986)). Perhaps the best studied of these are the Integrase/att system from bacteriophage λ (Landy, A. (1993) *Current Opinions in Genetics and Devel.* 3:699–707), the Cre/loxP system from bacteriophage P1 (Hoess and Abremski (1990) In *Nucleic Acids and Molecular Biology*, vol. 4. Eds.: Eckstein and Lilley, Berlin-Heidelberg: Springer-Verlag; pp. 90–109), and the FLP/FRT system from the Saccharomyces cerevisiae 2 μ circle plasmid (Broach et al. *Cell* 29:227–234 (1982)).

Members of a second family of site-specific recombinases, the resolvase family (e.g, γδ, Tn3 resolvase, Hin, Gin, and Cin) are also known. Members of this highly related family of recombinases are typically constrained to intramolecular reactions (e.g., inversions and excisions) and can require host-encoded factors. Mutants have been isolated that relieve some of the requirements for host factors (Maeser and Kahnmann (1991) *Mol. Gen. Genet.* 230:170–176), as well as some of the constraints of intramolecular recombination.

Other site-specific recombinases similar to λ Int and similar to P1Cre can be substituted for Int and Cre. Such recombinases are known. In many cases the purification of such other recombinases has been described in the art. In cases when they are not known, cell extracts can be used or the enzymes can be partially purified using procedures described for Cre and Int.

While Cre and Int are described in detail for reasons of example, many related recombinase systems exist and their application to the described invention is also provided according to the present invention. The integrase family of site-specific recombinases can be used to provide alternative recombination proteins and recombination sites for the present invention, as site-specific recombination proteins encoded by bacteriophage lambda, phi 80, P22, P2, 186, P4 and P1. This group of proteins exhibits an unexpectedly large diversity of sequences. Despite this diversity, all of the recombinases can be aligned in their C-terminal halves.

A 40-residue region near the C terminus is particularly well conserved in all the proteins and is homologous to a region near the C terminus of the yeast 2 mu plasmid Flp protein. Three positions are perfectly conserved within this family: histidine, arginine and tyrosine are found at respective alignment positions 396, 399 and 433 within the well-conserved C-terminal region. These residues contribute to the active site of this family of recombinases, and suggest that tyrosine-433 forms a transient covalent linkage to DNA during strand cleavage and rejoining. See, e.g, Argos, P. et al., *EMBO J.* 5:433–40 (1986).

Alternatively, IS231 and other Bacillus thuringiensis transposable elements could be used as recombination proteins and recombination sites. Bacillus thuringiensis is an entomopathogenic bacterium whose toxicity is due to the presence in the sporangia of delta-endotoxin crystals active against agricultural pests and vectors of human and animal diseases. Most of the genes coding for these toxin proteins are plasmid-borne and are generally structurally associated with insertion sequences (IS231, IS232, IS240, ISBT1 and ISBT2) and transposons (Tn4430 and Tn5401). Several of these mobile elements have been shown to be active and participate in the crystal gene mobility, thereby contributing to the variation of bacterial toxicity.

Structural analysis of the iso-IS231 elements indicates that they are related to IS1151 from Clostridium perfringens and distantly related to IS4 and IS186 from *Escherichia coli*. Like the other IS4 family members, they contain a conserved transposase-integrase motif found in other IS families and retroviruses.

Moreover, functional data gathered from IS231A in *Escherichia coli* indicate a non-replicative mode of transposition, with a preference for specific targets. Similar results were also obtained in *Bacillus subtilis* and *B. thuringiensis*. See, e.g., Mahillon, J. et al., *Genetica* 93:13–26 (1994); Campbell, *J. Bacteriol.* 7495–7499 (1992).

The amount of recombinase which is added to drive the recombination reaction can be determined by using known assays. Specifically, titration assay is used to determine the appropriate amount of a purified recombinase enzyme, or the appropriate amount of an extract.

Engineered Recombination Sites. The above recombinases and corresponding recombinase sites are suitable for use in recombination cloning according to the present invention. However, wild-type recombination sites contain sequences that reduce the efficiency or specificity of recombination reactions as applied in methods of the present invention. For example, multiple stop codons in attB, attR, attP, attL and loxP recombination sites occur in multiple reading frames on both strands, so recombination efficiencies are reduced, e.g., where the coding sequence must cross the recombination sites, (only one reading frame is available on each strand of loxP and attB sites) or impossible (in attP, attR or attL).

Accordingly, the present invention also provides engineered recombination sites that overcome these problems. For example, att sites can be engineered to have one or multiple mutations to enhance specificity or efficiency of the recombination reaction and the properties of Product DNAs (e.g, att1, att2, att3 sites); to decrease reverse reaction (e.g., removing P1and H1 from attB). The testing of these mutants determines which mutants yield sufficient recombinational activity to be suitable for recombination subcloning according to the present invention.

Mutations can therefore be introduced into recombination sites for enhancing site specific recombination. Such mutations include, but are not limited to: recombination sites without translation stop codons that allow fusion proteins to be encoded; recombination sites recognized by the same proteins but differing in base sequence such that they react largely or exclusively with their homologous partners allow multiple reactions to be contemplated. Which particular reactions take place can be specified by which particular partners are present in the reaction mixture. For example, a tripartite protein fusion could be accomplished with parental plasmids containing recombination sites attR1 and attR2; attL1 and attL3; and/or attR3 and attL2.

There are well known procedures for introducing specific mutations into nucleic acid sequences. A number of these are described in Ausubel, F. M. et al., *Current Protocols in Molecular Biology*, Wiley Interscience, New York (1989–1996). Mutations can be designed into oligonucleotides, which can be used to modify existing

cloned sequences, or in amplification reactions. Random mutagenesis can also be employed if appropriate selection methods are available to isolate the desired mutant DNA or RNA. The presence of the desired mutations can be confirmed by sequencing the nucleic acid by well known methods.

The following non-limiting methods can be used to engineer a core region of a given recombination site to provide mutated sites suitable for use in the present invention:

1. By recombination of two parental DNA sequences by site-specific (e.g. attL and attR to give attB) or other (e.g. homologous) recombination mechanisms. The DNA parental DNA segments containing one or more base alterations resulting in the final core sequence;
2. By mutation or mutagenesis (site-specific, PCR, random, spontaneous, etc) directly of the desired core sequence;
3. By mutagenesis (site-specific, PCR, random, spontanteous, etc) of parental DNA sequences, which are recombined to generate a desired core sequence; and
4. By reverse transcription of an RNA encoding the desired core sequence.

The functionality of the mutant recombination sites can be demonstrated in ways that depend on the particular characteristic that is desired. For example, the lack of translation stop codons in a recombination site can be demonstrated by expressing the appropriate fusion proteins. Specificity of recombination between homologous partners can be demonstrated by introducing the appropriate molecules into in vitro reactions, and assaying for recombination products as described herein or known in the art. Other desired mutations in recombination sites might include the presence or absence of restriction sites, translation or transcription start signals, protein binding sites, and other known functionalities of nucleic acid base sequences. Genetic selection schemes for particular functional attributes in the recombination sites can be used according to known method steps. For example, the modification of sites to provide (from a pair of sites that do not interact) partners that do interact could be achieved by requiring deletion, via recombination between the sites, of a DNA sequence encoding a toxic substance. Similarly, selection for sites that remove translation stop sequences, the presence or absence of protein binding sites, etc., can be easily devised by those skilled in the art.

Accordingly, the present invention provides a nucleic acid molecule, comprising at least one DNA segment having at least two engineered recombination sites flanking a Selectable marker and/or a desired DNA segment, wherein at least one of said recombination sites comprises a core region having at least one engineered mutation that enhances recombination in vitro in the formation of a Cointegrate DNA or a Product DNA.

The nucleic acid molecule can have at least one mutation that confers at least one enhancement of said recombination, said enhancement selected from the group consisting of substantially (i) favoring excisive integration; (ii) favoring excisive recombination; (ii) relieving the requirement for host factors; (iii) increasing the efficiency of said Cointegrate DNA or Product DNA formation; and (iv) increasing the specificity of said Cointegrate DNA or Product DNA formation.

The nucleic acid molecule preferably comprises at least one recombination site derived from attB, attP, attL or attR. More preferably the att site is selected from att1, att2, or att3, as described herein.

In a preferred embodiment, the core region comprises a DNA sequence selected from the group consisting of:

(a) RKYCWGCTTFYKTRTACNAASTSGB (m-att) (SEQ ID NO:1);

(b) AGCCWGCTTTYKTRTACNAACTSGB (m-attB) (SEQ ID NO:2);

(c) GTTCAGCTTTCKTRTACNAACTSGB (m-attR) (SEQ ID NO:3);

(d) AGCCWGCTTTCKTRTACNAAGTSGB (m-attL) (SEQ ID NO:4);

(e) GTTCAGCTTTYKTRTACNAAGTSGB(m-attP1) (SEQ ID NO:5);

or a corresponding or complementary DNA or RNA sequence, wherein R=A or G; K=G or T/U; Y=C or T/U; W=A or T/U; N=A or C or G or T/U; S=Cor G; and B=C or G or T/U, as presented in 37 C.F.R. §1.822, which is entirely incorporated herein by reference, wherein the core region does not contain a stop codon in one or more reading frames.

The core region also preferably comprises a DNA sequence selected from the group consisting of:

(a) AGCCTGCTTTTTTGTACAAACTTGT(attB1) (SEQ ID NO:6);

(b) AGCCTGCTTTCTTGTACAAACTTGT (attB2) (SEQ ID NO:7);

(c) ACCCAGCTFTCTTGTACAAACTTGT (attB3) (SEQ ID NO:8);

(d) GTTCAGCTTTGTACAAACTTGT (attR1) (SEQ ID NO:9);

(e) GTTCAGCTTRCTTGTACAAACTTGT (attnR2) (SEQ ID NO:10);

(f) GTTCAGCTTTCTTGTACAAAGTTGG (attR3) (SEQ ID NO:11);

(g) AGCCTGCTTTTTTGTACAAAGTTGG (attL1) (SEQ ID NO:12);

(h) AGCCTGCTTTCTTGTACAAAGTTGG (attL2) (SEQ ID NO:13);

(i) ACCCAGCTTTCTTGTACAAAGTTGG (attL3) (SEQ ID NO:14);

(j) GTTCAGCTTTTTTGTACAAAGTTGG(attP1) (SEQ ID NO: 15);

(k) GTTCAGCTTTCTTGTACAAAGTTGG (attP2,P3) (SEQ ID NO:16); or a corresponding or complementary DNA or RNA sequence.

The present invention thus also provides a method for making a nucleic acid molecule, comprising providing a nucleic acid molecule having at least one engineered recombination site comprising at least one DNA sequence having at least 80–99% homology (or any range or value therein) to at least one of SEQ ID NOS:1–16, or any suitable recombination site, or which hybridizes under stringent conditions thereto, as known in the art.

Clearly, there are various types and permutations of such well-known in vitro and in vivo selection methods, each of which are not described herein for the sake of brevity. However, such variations and permutations are contemplated and considered to be the different embodiments of the present invention.

It is important to note that as a result of the preferred embodiment being in vitro recombination reactions, non-biological molecules such as PCR products can be manipulated via the present recombinational cloning method. In one example, it is possible to clone linear molecules into circular vectors. There are a number of applications for the present invention. These uses include, but are not limited to, chang-

ing vectors, apposing promoters with genes, constructing genes for fusion proteins, changing copy number, changing replicons, cloning into phages, and cloning, e.g, PCR products (with an attB site at one end and a loxP site at the other end), genomic DNAs, and cDNAs.

The following examples are intended to further illustrate certain preferred embodiments of the invention and are not intended to be limiting in nature.

EXAMPLES

The present recombinational cloning method accomplishes the exchange of nucleic acid segments to render something useful to the user, such as a change of cloning vectors. These segments must be flanked on both sides by recombination signals that are in the proper orientation with respect to one another. In the examples below the two parental nucleic acid molecules (e.g., plasmids) are called the Insert Donor and the Vector Donor. The Insert Donor contains a segment that will become joined to a new vector contributed by the Vector Donor. The recombination intermediate(s) that contain(s) both starting molecules is called the Cointegrate(s). The second recombination event produces two daughter molecules, called the Product (the desired new clone) and the Byproduct.

Buffers

Various known buffers can be used in the reactions of the present invention. For restriction enzymes, it is advisable to use the buffers recommended by the manufacturer. Alternative buffers can be readily found in the literature or can be devised by those of ordinary skill in the art.

Examples 1–3. One exemplary buffer for lambda integrase is comprised of 50 mM Tris-HCl, at pH 7.5–7.8, 70 mM KCl, 5 mM spermidine, 0.5 mM EDTA, and 0.25 mg/ml bovine serum albumin, and optionally, 10% glycerol.

One preferred buffer for P1 Cre recombinase is comprised of 50 mM Tris-HCl at pH 7.5, 33 mM NaCl, 5 mM spermidine, and 0.5 mg/ml bovine serum albumin.

The buffer for other site-specific recombinases which are similar to lambda Int and P1 Cre are either known in the art or can be determined empirically by the skilled artisans, particularly in light of the above-described buffers.

Example 1

Recombinational Cloning Using Cre and Cre & Int

Two pairs of plasmids were constructed to do the in vitro recombinational cloning method in two different ways. One pair, pEZC705 and pEZC726 (FIG. 2A), was constructed with loxP and att sites, to be used with Cre and λ integrase. The other pair, pEZC602 and pEZC629 (FIG. 3A), contained the loxP (wild type) site for Cre, and a second mutant lox site, loxP 511, which differs from loxP in one base (out of 34 total). The minimum requirement for recombinational cloning of the present invention is two recombination sites in each plasmid, in general X and Y, and X'and Y'. Recombinational cloning takes place if either or both types of site can recombine to form a Cointegrate (e.g. X and X'), and if either or both (but necessarily a site different from the type forming the Cointegrate) can recombine to excise the Product and Byproduct plasmids from the Cointegrate (e.g. Y and Y'). It is important that the recombination sites on the same plasmid do not recombine. It was found that the present recombinational cloning could be done with Cre alone.

Cre-Only

Two plasmids were constructed to demonstrate this conception (see FIG. 3A). pEZC629 was the Vector Donor plasmid. It contained a constitutive drug marker (chloramphenicol resistance), an origin of replication, loxP and loxP 511 sites, a conditional drug marker (kanamycin resistance whose expression is controlled by the operator/promoter of the tetracycline resistance operon of transposon Tn10), and a constitutively expressed gene for the tet repressor protein, tetR. *E. coli* cells containing pEZC629 were resistant to chloramphenicol at 30 μg/ml, but sensitive to kanamycin at 100 μg/ml. pEZC602 was the Insert Donor plasmid, which contained a different drug marker (ampicillin resistance), an origin, and loxP and loxP 511 sites flanking a multiple cloning site.

This experiment was comprised of two parts as follows:

Part I: About 75 ng each of pEZC602 and pEZC629 were mixed in a total volume of 30 μl of Cre buffer (50 mM Tris-HCl pH 7.5, 33 mM NaCl, 5 mM spermidine-HCl, 500 μg/ml bovine serum albumin). Two 10 μl aliquots were transferred to new tubes. One tube received 0.5 μl of Cre protein (approx. 4 units per μl; partially purified according to Abremski and Hoess, *J. Biol. Chem.* 259:1509 (1984)). Both tubes were incubated at 37° C. for 30 minutes, then 70° C. for 10 minutes. Aliquots of each reaction were diluted and transformed into DH5α. Following expression, aliquots were plated on 30 μg/ml chloramphenicol; 100 μg/ml ampicillin plus 200 μg/ml methicillin; or 100 μg/ml kanamycin. Results: See Table 1. The reaction without Cre gave 1.11×10^6 ampicillin resistant colonies (from the Insert Donor plasmid pEZC602); 7.8×10^5 chloramphenicol resistant colonies (from the Vector Donor plasmid pEZC629); and 140 kanamycin resistant colonies (background). The reaction with added Cre gave 7.5×10^5 ampicillin resistant colonies (from the Insert Donor plasmid pEZC602); 6.1×10^5 chloramphenicol resistant colonies (from the Vector Donor plasmid pEZC629); and 760 kanamycin resistant colonies (mixture of background colonies and colonies from the recombinational cloning Product plasmid). Analysis: Because the number of colonies on the kanamycin plates was much higher in the presence of Cre, many or most of them were predicted to contain the desired Product plasmid.

TABLE 1

Enzyme	Ampicillin	Chloramphenicol	Kanamycin	Efficiency
None	1.1×10^6	7.8×10^5	140	140/7.8 × 10^5 = 0.02%
Cre	7.5×10^5	6.1×10^5	760	760/6.1 × 10^5 = 0.12%

Part II: Twenty four colonies from the "+Cre" kanamycin plates were picked and inoculated into medium containing 100 μg/ml kanamycin. Minipreps were done, and the miniprep DNAs, uncut or cut with Smalor HindIII, were electrophoresed. Results: 19 of the 24 minipreps showed supercoiled plasmid of the size predicted for the Product plasmid. All 19 showed the predicted SmaI and HindIII restriction fragments. Analysis: The Cre only scheme was demonstrated. Specifically, it was determined to have yielded about 70% (19 of 24) Product clones. The efficiency was about 0.1% (760 kanamycin resistant clones resulted from 6.1×10^5 chloramphenicol resistant colonies).

Cre Plus Integrase

The plasmids used to demonstrate this method are exactly analogous to those used above, except that pEZC726, the Vector Donor plasmid, contained an attP site in place of loxP 511, and pEZC705, the Insert Donor plasmid, contained an attB site in place of loxP 511 (FIG. 2A).

This experiment was comprised of three parts as follows:

Part I: About 500 ng of pEZC705 (the Insert Donor plasmid) was cut with ScaI, which linearized the plasmid

within the ampicillin resistance gene. (This was done because the λ integrase reaction has been historically done with the attB plasmid in a linear state (H. Nash, personal communication). However, it was found later that the integrase reaction proceeds well with both plasmids supercoiled.) Then, the linear plasmid was ethanol precipitated and dissolved in 20 μl of λ integrase buffer (50 mM Tris-HCl, about pH 7.8, 70 mM KCl, 5 mM spermidine-HCl, 0.5 mM EDTA, 250 μg/ml bovine serum albumin). Also, about 500 ng of the Vector Donor plasmid pEZC726 was ethanol precipitated and dissolved in 20 μl λ integrase buffer. Just before use, λ integrase (2 μl, 393 μg/ml) was thawed and diluted by adding 18 μl cold λ integrase buffer. One μl IHF (integration host factor, 2.4 mg/ml, an accessory protein) was diluted into 150 μl cold λ integrase buffer. Aliquots (2 μl) of each DNA were mixed with λ integrase buffer, with or without 1 μl each λ integrase and IHF, in a total of 10 μl. The mixture was incubated at 25° C. for 45 minutes, then at 70° C. for 10 minutes. Half of each reaction was applied to an agarose gel. Results: In the presence of integrase and IHF, about 5% of the total DNA was converted to a linear Cointegrate form. Analysis: Activity of integrase and IHF was confirmed.

Part II: Three microliters of each reaction (i.e., with or without integrase and IHF) were diluted into 27 μl of Cre buffer (above), then each reaction was split into two 10 μl aliquots (four altogether). To two of these reactions, 0.5 μl of Cre protein (above) were added, and all reactions were incubated at 37° C. for 30 minutes, then at 70° C. for 10 minutes. TE buffer (90 μl; TE: 10 mM Tris-HCl, pH 7.5, 1 mM EDTA) was added to each reaction, and 1 μl each was transformed into *E. coli* DH5α. The transformation mixtures were plated on 100 μg/ml ampicillin plus 200 μg/ml methicillin; 30 μg/ml chloramphenicol; or 100 μg/ml kanamycin. Results: See Table 2.

TABLE 2

Enzyme	Ampicillin	Chloram-phenicol	Kanamycin	Efficiency
None	990	20000	4	$4/2 \times 10^4 = 0.02\%$
Cre only	280	3640	0	0
Integrase* only	1040	27000	9	$9/2.7 \times 10^4 = 0.03\%$
Integrase* + Cre	110	1110	76	$76/1.1 \times 10^3 = 6.9\%$

*Integrase reactions also contained IHF.

Analysis: The Cre protein impaired transformation. When adjusted for this effect, the number of kanamycin resistant colonies, compared to the control reactions, increased more than 100 fold when both Cre and Integrase were used. This suggests a specificity of greater than 99%.

Part III: 38 colonies were picked from the Integrase plus Cre plates, miniprep DNAs were made and cut with HindIII to give diagnostic mapping information. Result: All 38 had precisely the expected fragment sizes. Analysis: The Cre plus λ integrase method was observed to have much higher specificity than Cre-alone. Conclusion: The Cre plus λ integrase method was demonstrated. Efficiency and specificity were much higher than for Cre only.

Example 2

Using in vitro Recombinational Cloning to Subclone the Chloramphenicol Acetyl Transferase Gene into a Vector for Expression in Eukaryotic Cells (FIG. 4A)

An Insert Donor plasmid, pEZC843, was constructed, comprising the chloramphenicol acetyl transferase gene of *E. coli*, cloned between loxP and attB sites such that the loxP site was positioned at the 5'-end of the gene (FIG. 4B). A Vector Donor plasmid, pEZC1003, was constructed, which contained the cytomegalovirus eukaryotic promoter apposed to a loxP site (FIG. 4C). One microliter aliquots of each supercoiled plasmid (about 50 ng crude miniprep DNA) were combined in a ten microliter reaction containing equal parts of lambda integrase buffer (50 mM Tris-HCl, pH 7.8, 70 mM KCl, 5 mM spermidine, 0.5 mM EDTA, 0.25 mg/ml bovine serum albumin) and Cre recombinase buffer (50 mM Tris-HCl, pH 7.5, 33 mM NaCl, 5 mM spermidine, 0.5 mg/ml bovine serum albumin), two units of Cre recombinase, 16 ng integration host factor, and 32 ng lambda integrase. After incubation at 30° C. for 30 minutes and 75° C. for 10 minutes, one microliter was transformed into competent *E. coli* strain DH5α (Life Technologies, Inc.). Aliquots of transformations were spread on agar plates containing 200 μg/ml kanamycin and incubated at 37° C. overnight. An otherwise identical control reaction contained the Vector Donor plasmid only. The plate receiving 10% of the control reaction transformation gave one colony; the plate receiving 10% of the recombinational cloning reaction gave 144 colonies. These numbers suggested that greater than 99% of the recombinational cloning colonies contained the desired product plasmid. Miniprep DNA made from six recombinational cloning colonies gave the predicted size plasmid (5026 base pairs), CMVProd. Restriction digestion with NcoI gave the fragments predicted for the chloramphenicol acetyl transferase cloned downstream of the CMV promoter for all six plasmids.

Example 3

Subcloned DNA Segments Flanked by attB Sites Without Stop Codons

Part I: Background

The above examples are suitable for transcriptional fusions, in which transcription crosses recombination sites. However, both attR and loxP sites contain multiple stop codons on both strands, so translational fusions can be difficult, where the coding sequence must cross the recombination sites, (only one reading frame is available on each strand of loxP sites) or impossible (in attR or attL).

A principal reason for subcloning is to fuse protein domains. For example, fusion of the glutathione S-transferase (GST) domain to a protein of interest allows the fusion protein to be purified by affinity chromatography on glutathione agarose (Pharmacia, Inc., 1995 catalog). If the protein of interest is fused to runs of consecutive histidines (for example His6), the fusion protein can be purified by affinity chromatography on chelating resins containing metal ions (Qiagen, Inc.). It is often desirable to compare amino terminal and carboxy terminal fusions for activity, solubility, stability, and the like.

The attB sites of the bacteriophage λ integration system were examined as an alternative to loxP sites, because they are small (25 bp) and have some sequence flexibility (Nash, H. A. et al., *Proc. Natl. Acad. Sci. USA* 84:4049–4053 (1987). It was not previously suggested that multiple muta-

tions to remove all stop codes would result in useful recombination sites for recombinational subcloning.

Using standard nomenclature for site specific recombination in lambda bacteriophage (Weisber, in *Lambda III*, Hendrix, et al., eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1989)), the nucleotide regions that participate in the recombination reaction in an *E. coli* host cell are represented as follows:

attP —P1—H1—P2—X—H2—C—O—C—H'—P'1—P'2—P'3—

+

attB —B—O—B'—

Int, IHF ↓↑ Xis, Int, IHF

attR —P1—H1—P2—X—H2—C—O—B'—

+

attL —B—O—C—H'—P'1—P'2—P'3—,

where: O represents the 15 bp core DNA sequence found in both the phage and *E. coli* genomes; B and B' represent approximately 5 bases adjacent to the core in the *E. coli* genome; and P1, H1, P2, X, H2, C, C', H', P'1, P'2, and P'3 represent known DNA sequences encoding protein binding domains in the bacteriophage λ genome.

The reaction is reversible in the presence of the protein Xis (excisionase); recombination between attL, and attR precisely excise the λ genome from its integrated state, regenerating the circular λ genome containing attp and the linear *E. coli* genome containing attB.

Part II: Construction and Testing of Plasmids Containing Mutant att Sites

Mutant attL and attR sites were constructed. Importantly, Landy et al. (*Ann. Rev. Biochem.* 58:913 (1989)) observed that deletion of the P1and H1 domains of attP facilitated the excision reaction and eliminated the integration reaction, thereby making the excision reaction irreversible. Therefore, as mutations were introduced in attR, the P1and H1 domains were also deleted. attR sites in the present example lack the P1and H1 regions and have the NdeI site removed (base 27630 changed from C to G), and contain sequences corresponding to bacteriophage λ coordinates 27619–27738 (GenBank release 92.0, bg:LAMCG, "Complete Sequence of Bacteriophage Lambda").

The sequence of attB produced by recombination of wild type attL and attR sites is:

		B	O	B'	
attBwt:	5'	AGCCT	GCTTTTTTATACTAA	CTTGA	3' (SEQ. ID NO:31)
	3'	TCGGA	CGAAAAAATATGATT	GAACT	5' (SEQ. ID NO:32)

The stop codons are italicized and underlined. Note that sequences of attL, attR, and attP can be derived from the attB sequence and the boundaries of bacteriophage λ contained within attL and attR (coordinates 27619 to 27818).

When mutant attR1 and attL1 sites were recombined the sequence attB1 was produced (mutations in bold, large font):

		B	O	B'	
attB1:	5'	AGCCT	GCTTTTTTGTACAAA CTTGT	3	(SEQ. ID NO:6)
	3'	TCGGA	CGAAAAAACATGTTT GAACA	5'	(SEQ. ID NO:33)

Note that the four stop codons are gone.

When an additional mutation was introduced in the attR1 and attL1 sequences (bold), attR2 and attL2 sites resulted. Recombination of attR2 and attL2 produced the attB2 site:

		B	O	B'	
attB2:	5'	AGCCT	GCTTTCTTGTACAAA CTTGT	3'	(SEQ. ID NO:7)
	3'	TCGGA	CGAAAGAACATGTTT GAACA	5'	(SEQ. ID NO:34)

The recombination activities of the above attL and attR sites were assayed as follows. The attB site of plasmid pEZC705 (FIG. 2B) was replaced with attLwt, attL1, or attL2. The attP site of plasmid pEZC726 (FIG. 2C) was replaced with attRwt (lacking regions P1and H1), attR1, or attR2. Thus, the resulting plasmids could recombine via their loxP sites, mediated by Cre, and via their attR and attL sites, mediated by Int, Xis, and IHF. Pairs of plasmids were mixed and reacted with Cre, Int, Xis, and IHF, transformed into *E. coli* competent cells, and plated on agar containing kanamycin. The results are presented in Table 3:

TABLE 3

Vector donor att site	Gene donor att site	# of kanamycin resistant colonies*
attRwt (pEZC1301)	None	1 (background)
"	attLwt (pEZC1313)	147
"	attL1 (pEZC1317)	47
"	attL2 (pEZC1321)	0
attR1 (pEZC1305)	None	1 (background)
"	attLwt (pEZC1313)	4
"	attL1 (pEZC1317)	128
"	attL2 (pEZC1321)	0
attR2 (pEZC1309)	None	9 (background)
"	attLwt (pEZC1313)	0
"	attL2 (pEZC1317)	0
"	attL2 (pEZC1321)	209

(*1% of each transformation was spread on a kanamycin plate.)

The above data show that whereas the wild type att and attl sites recombine to a small extent, the attl and att2 sites do not recombine detectably with each other.

Part III. Recombination was demonstrated when the core region of both attb sites flanking the DNA segment of interest did not contain stop codons. The physical state of the participating plasmids was discovered to influence recombination efficiency.

The appropriate att sites were moved into pEZC705 and pEZC726 to make the plasmids pEZC1405 (FIG. 5G) (attR1 and attR2) and pEZC1502 (FIG. 5H) (attL1 and attL2). The desired DNA segment in this experiment was a copy of the chloramphenicol resistance gene cloned between the two attL sites of pEZC1502. Pairs of plasmids were recombined in vitro using Int, Xis, and IHF (no Cre because no loxP sites were present). The yield of desired kanamycin resistant colonies was determined when both parental plasmids were circular, or when one plasmid was circular and the other linear as presented in Table 4:

TABLE 4

Vector donor[1]	Gene donor[1]	Kanamycin resistant colonies[2]
Circular pEZC1405	None	30
Circular pEZC1405	Circular pEZC1502	2680
Linear pEZC1405	None	90
Linear pEZC1405	Circular pEZC1502	172000
Circular pEZC1405	Linear pEZC1502	73000

[1]DNAs were purified with Qiagen columns, concentrations determined by A260, and linearized with Xba I (pEZC1405) or AlwN I (pEZC1502). Each reaction contained 200 ng of the indicated DNA. All reactions (10 μl total) contained 3 μl of enzyme mix (Xis, Int, and IHF). After incubation (45 minutes at 25°, 10 minutes at 65°), one μl was used to transform *E. coli* DH5α cells.

[2]Number of colonies expected if the entire transformation reaction (1 ml) had been plated. Either 100 μl or 1 μl of the transformations were actually plated.

Analysis: Recombinational cloning using mutant attR and attL sites was confirmed. The desired DNA segment is subcloned between attB sites that do not contain any stop codons in either strand. The enhanced yield of Product DNA (when one parent was linear) was unexpected because of earlier observations that the excision reaction was more efficient when both participating molecules were supercoiled and proteins were limiting (Nunes-Duby et al., *Cell* 50:779–788 (1987).

Example 4

Demonstration of Recombinational Cloning Without Inverted Repeats

Part I. Rationale

The above Example 3 showed that plasmids containing inverted repeats of the appropriate recombination sites (for example, attL1 and attL2 in plasmid pEZC1502) (FIG. 5H) could recombine to give the desired DNA segment flanked by attB sites without stop codons, also in inverted orientation. A concern was the in vivo and in vitro influence of the inverted repeats. For example, transcription of a desired DNA segment flanked by attB sites in inverted orientation could yield a single stranded RNA molecule that might form a hairpin structure, thereby inhibiting translation.

Inverted orientation of similar recombination sites can be avoided by placing the sites in direct repeat arrangement att sites. If parental plasmids each have a wild type attL and wild type attR site, in direct repeat the Int, Xis, and IHF proteins will simply remove the DNA segment flanked by those sites in an intramolecular reaction. However, the mutant sites described in the above Example 3 suggested that it might be possible to inhibit the intramolecular reaction while allowing the intermolecular recombination to proceed as desired.

Part II: Structure of Plasmids Without Inverted Repeats for Recombinational Cloning

The attR2 sequence in plasmid pEZC1405 (FIG. 5G) was replaced with attL2, in the opposite orientation, to make pEZC1603 (FIG. 6A). The attL2 sequence of pEZC1502 (FIG. 5H) was replaced with attR2, in the opposite orientation, to make pEZC1706 (FIG. 6B). Each of these plasmids contained mutations in the core region that make intramolecular reactions between att1 and att2 cores very inefficient (see Example 3, above).

Plasmids pEZC1405, pEZC1502, pEZC1603 and pEZC1706 were purified on Qiagen columns (Qiagen, Inc.). Aliquots of plasmids pEZC1405 and pEZC1603 were linearized with Xba I. Aliquots of plasmids pEZC1502 and pEZC1706 were linearized with AlwN I. One hundred ng of plasmids were mixed in buffer (equal volumes of 50 mM Tris HCl pH 7.5, 25 mM Tris HCl pH 8.0, 70 mM KCl, 5 mM spermidine, 0.5 mM EDTA, 250 μg/ml BSA, 10% glycerol) containing Int (43.5 ng), Xis (4.3 ng) and IHF (8.1 ng) in a final volume of 10 μl. Reactions were incubated for 45 minutes at 25° C., 10 minutes at 65 ° C., and 1 μl was transformed into *E. coli* DH5α. After expression, aliquots were spread on agar plates containing 200 1μg/ml kanamycin and incubated at 37° C.

Results, expressed as the number of colonies per 1 μl of recombination reaction are presented in Table 5:

TABLE 5

Vector Donor	Gene Donor	Colonies	Predicted % product
Circular 1405	—	100	—
Circular 1405	Circular 1502	3740	3640/3740 = 97%
Linear 1405	—	90	—
Linear 1405	Circular 1502	172,000	171,910/172,000 = 99.9%
Circular 1405	Linear 1502	73,000	72,900/73,000 = 99.9%
Circular 1603	—	80	—
Circular 1603	Circular 1706	410	330/410 = 80%
Linear 1603	—	270	—
Linear 1603	Circular 1706	7000	6730/7000 = 96%
Circular 1603	Linear 1706	10,800	10,530/10,800 = 97%

Analysis. In all configurations, i.e., circular or linear, the pEZC1405×pEZC1502 pair (with att sites in inverted repeat configuration) was more efficient than pEZC1603×

pEZC1706 pair (with att sites mutated to avoid hairpin formation). The pEZC1603×pEZC1706 pair gave higher backgrounds and lower efficiencies than the pEZC1405× pEZC1502 pair. While less efficient, 80% or more of the colonies from the pEZC1603×pEZC1706 reactions were expected to contain the desired plasmid product. Making one partner linear stimulated the reactions in all cases.

Part III: Confirmation of Product Plasmids' Structure

Six colonies each from the linear pEZC1405 (FIG. 5G)× circular pEZC1502 (FIG. 5H), circular pEZC1405×linear pEZC1502, linear pEZC1603 (FIG. 6A)×circular pEZC1706 (FIG. 6B), and circular pEZC1603×linear pEZC1706 reactions were picked into rich medium and miniprep DNAs were prepared. Diagnostic cuts with Ssp I gave the predicted restriction fragments for all 24 colonies.

Analysis. Recombination reactions between plasmids with mutant attL and attR sites on the same molecules gave the desired plasmid products with a high degree of specificity.

Example 5

Recombinational Cloning with a Toxic Gene

Part I: Background

Restriction enzyme Dpn I recognizes the sequence GATC and cuts that sequence only if the A is methylated by the dam methylase. Most commonly used E. co/i strains are dam+. Expression of Dpn I in dam+ strains of *E. coli* is lethal because the chromosome of the cell is chopped into many pieces. However, in dam- cells expression of Dpn I is innocuous because the chromosome is immune to Dpn I cutting.

In the general recombinational cloning scheme, in which the vector donor contains two segments C and D separated by recombination sites, selection for the desired product depends upon selection for the presence of segment D, and the absence of segment C. In the original Example segment D contained a drug resistance gene (Km) that was negatively controlled by a repressor gene found on segment C. When C was present, cells containing D were not resistant to kanamycin because the resistance gene was turned off.

The Dpn I gene is an example of a toxic gene that can replace the repressor gene of the above embodiment. If segment C expresses the Dpn I gene product, transforming plasmid CD into a dam+ host kills the cell. If segment D is transferred to a new plasmid, for example by recombinational cloning, then selecting for the drug marker will be successful because the toxic gene is no longer present.

Part II: Construction of a Vector Donor Using Dpn I as a Toxic Gene

The gene encoding Dpn I endonuclease was amplified by PCR using primers 5'CCA CCA CAAACG CGT CCA TGG AAT TAC ACT TTA ATT TAG3' (SEQ. ID NO: 17) and 5'CCA CCA CAA GTC GAC GCA TGC CGA CAG CCT TCC AAA TGT3' (SEQ. ID NO:18) and a plasmid containing the Dpn I gene (derived from plasmids obtained from Sanford A. Lacks, Brookhaven National Laboratory, Upton, N.Y.; also available from American Type Culture Collection as ATCC 67494) as the template.

Additional mutations were introduced into the B and B' regions of attL and attR, respectively, by amplifying existing attL and attR domains with primers containing the desired base changes. Recombination of the mutant attL3 (made with oligo Xis115) and attR3 (made with oligo Xis112) yielded attB3 with the following sequence (differences from attB1 in bold):

B	O	B'
5'ACCCA	GCTTTCTTGTACAAA	GTGGT 3' (SEQ. ID NO:8)
3'TGGGT	CGAAAGAACATGTTT	CACCA 5' (SEQ. ID NO:35)

The attL3 sequence was cloned in place of attL2 of an existing Gene Donor plasmid to give the plasmid pEZC2901 (FIG. 7A). The attR3 sequence was cloned in place of attR2 in an existing Vector Donor plasmid to give plasmid pEZC2913 (FIG. 7B) Dpn I gene was cloned into plasmid pEZC2913 to replace the tet repressor gene. The resulting Vector Donor plasmid was named pEZC3101 (FIG. 7C). When pEZC3101 was transformed into the dam- strain SCS110 (Stratagene), hundreds of colonies resulted. When the same plasmid was transformed into the dam+ strain DH5α, only one colony was produced, even though the DH5α cells were about 20 fold more competent than the SCS110 cells. When a related plasmid that did not contain the Dpn I gene was transformed into the same two cell lines, 28 colonies were produced from the SCS110 cells, while 448 colonies resulted from the DH5α cells. This is evidence that the Dpn I gene is being expressed on plasmid pEZC3101 (FIG. 7C), and that it is killing the dam+ DH5α cells but not the dam- SCS110 cells.

Part III: Demonstration of Recombinational Cloning Using Dpn I Selection

A pair of plasmids was used to demonstrate recombinational cloning with selection for product dependent upon the toxic gene Dpn 1. Plasmid pEZC3101 (FIG. 7C) was linearized with Mlu I and reacted with circular plasmid pEZC2901 (FIG. 7A). A second pair of plasmids using selection based on control of drug resistance by a repressor gene was used as a control: plasmid pEZC1802 (FIG. 7D) was linearized with Xba I and reacted with circular plasmid pEZC1502 (FIG. 5H). Eight microliter reactions containing the same buffer and proteins Xis, Int, and IHF as in previous examples were incubated for 45 minutes at 25° C., then 10 minutes at 75° C., and 1 μl aliquots were transformed into DH5α (i.e., dam+) competent cells, as presented in Table 6.

TABLE 6

Reaction #	Vector donor	Basis of selection	Gene donor	Colonies
1	pEZC3101/Mlu	Dpn I toxicity	—	3
2	pEZC3101/Mlu	Dpn I toxicity	Circular pEZC2901	4000
3	pEZC1802/Xba	Tet repressor	—	0
4	pEZC1802/Xba	Tet repressor	Circular pEZC1502	12100

Miniprep DNAs were prepared from four colonies from reaction #2, and cut with restriction enzyme Ssp I. All gave the predicted fragments.

Analysis: Subcloning using selection with a toxic gene was demonstrated. Plasmids of the predicted structure were produced.

Example 6

Cloning of Genes with UraciI DNA Glycosylase and Subcloning of the Genes with Recombinational Cloning to Make Fusion Proteins

Part I: Converting an Existing Expression Vector to a Vector Donor for Recombinational Cloning

A cassette useful for converting existing vectors into functional Vector Donors was made as follows. Plasmid pEZC3101 (FIG. 7C) was digested with Apa I and Kpn I, treated with T4 DNA polymerase and dNTPs to render the

ends blunt, further digested with Sma I, Hpa I, and AlwN I to render the undesirable DNA fragments small, and the 2.6 kb cassette containing the attR1-Cm^R-Dpn I-attR-3 domains was gel purified. The concentration of the purified cassette was estimated to be about 75 ng DNA/μl.

Plasmid pGEX-2TK (FIG. 8A) (Pharmacia) allows fusions between the protein glutathione S transferase and any second coding sequence that can be inserted in its multiple cloning site. pGEX-2TK DNA was digested with Sma I and treated with alkaline phosphatase. About 75 ng of the above purified DNA cassette was ligated with about 100 ng of the pGEX-2TK vector for 2.5 hours in a 5 μl ligation, then 1 μl was transformed into competent BRL 3056 cells (a dam$^-$ derivative of DH10B; dam$^-$ strains commercially available include DM1 from Life Technologies, Inc., and SCS 110 from Stratagene). Aliquots of the transformation mixture were plated on LB agar containing 100 μg/ml ampicillin (resistance gene present on pGEX-2TK) and 30 μg/ml chloramphenicol (resistance gene present on the DNA cassette). Colonies were picked and miniprep DNAs were made. The orientation of the cassette in pGEX-2TK was determined by diagnostic cuts with EcoR I. A plasmid with the desired orientation was named pEZCC3501 (FIG. 8B).

Part II: Cloning Reporter Genes Into an Recombinational Cloning Gene Donor Plasmid in Three Reading Frames

Uracil DNA glycosylase (UDG) cloning is a method for cloning PCR amplification products into cloning vectors (U.S. Pat. No. 5,334,515, entirely incorporated herein by reference). Briefly, PCR amplification of the desired DNA segment is performed with primers that contain uracil bases in place of thymidine bases in their 5' ends. When such PCR products are incubated with the enzyme UDG, the uracil bases are specifically removed. The loss of these bases weakens base pairing in the ends of the PCR product DNA, and when incubated at a suitable temperature (e.g., 37° C.), the ends of such products are largely single stranded. If such incubations are done in the presence of linear cloning vectors containing protruding 3' tails that are complementary to the 3' ends of the PCR products, base pairing efficiently anneals the PCR products to the cloning vector. When the annealed product is introduced into *E. coli* cells by transformation, in vivo processes efficiently convert it into a recombinant plasmid.

UDG cloning vectors that enable cloning of any PCR product in all three reading frames were prepared from pEZC3201 (FIG. 8K) as follows. Eight oligonucleotides were obtained from Life Technologies, Inc. (all written 5'→3': rf1 top (GGCC GAT TAC GAT ATC CCA ACG ACC GAA AAC CTG TAT TTT CAG GCT) (SEQ. ID NO:19), rf1 bottom (CAG GTT,RTC GGT CGT TGG GAT ATC GTA ATC)(SEQ. ID NO:20), rf2 top (GGCCA GAT TAC GAT ATC CCA ACG ACC GAA AAC CTG TAT TTT CAG GGT)(SEQ. ID NO:21), rf2 bottom (CAG GTT TTC GGT CCT TGG GAT ATC GTA ATC TY(SEQ. ID NO:22), rf3 top (GGCCAA GAT TAC GAT ATC CCA ACG ACC GAA AAC CTG TAT TTT CAG GGT)(SEQ. ID NO:23), rf3 bottom (CAGCGTT TTC GGT CGT TGG GAT ATC GTA ATC TT)(SEQ. ID NO:24), carboxy top (ACC GTT TAC GTC GAC)(SEQ. ID NO:25) and carboxy bottom (TCGA GTC CAC GTA AAC GGT TCC CAC TTA TTA)(SEQ. ID NO:26). The rf1, 2, and 3 top strands and the carboxy bottom strand were phosphorylated on their 5' ends with T4 polynucleotide kinase, and then the complementary strands of each pair were hybridized. Plasmid pEZC3201 (FIG. 8K) was cut with Not I and Sal I, and aliquots of cut plasmid were mixed with the carboxy-oligo duplex (Sal I end) and either the rf1, rf2, or rf3 duplexes (Not I ends) (10 μg cut plasmid (about 5 pmol) mixed with 250 pmol carboxy oligo duplex, split into three 20 μl volumes, added 5 μl (250 pmol) of rf1, rf2, or rf3 duplex and 2 μl=2 units T4 DNA ligase to each reaction). After 90 minutes of ligation at room temperature, each reaction was applied to a preparative agarose gel and the 2.1 kb vector bands were eluted and dissolved in 50 μl of TE.

Part III: PCR of CAT and phoA Genes

Primers were obtained from Life Technologies, Inc., to amplify the chloramphenicol acetyl transferase (CAT) gene from plasmid pACYC184, and phoA, the alkaline phosphatase gene from *E. coli*. The primers had 12-base 5' extensions containing uracil bases, so that treatment of PCR products with uracil DNA glycosylase (UDG) would weaken base pairing at each end of the DNAs and allow the 3' strands to anneal with the protruding 3' ends of the rf1, 2, and 3 vectors described above. The sequences of the primers (all written 5'→3') were: CAT left, UAU UUU CAG GGU ATG GAG AAA AAA ATC ACT GGA TAT ACC (SEQ. ID NO:27); CAT right, UCC CAC UUA UUA CGC CCC GCC CTG CCA CTC ATC (SEQ. ID NO:28); phoA left, UAU UUU CAG GGU ATG CCT GTT CTG GAA AAC CGG (SEQ. ID NO:29); and phoA right, UCC CAC UUA UUA TTT CAG CCC CAG GGC GGC TTT C (SEQ. ID NO:30). The primers were then used for PCR reactions using known method steps (see, e.g., U.S. Pat. No. 5,334,515, entirely incorporated herein by reference), and the polymerase chain reaction amplification products obtained with these primers comprised the CAT or phoA genes with the initiating ATGs but without any transcriptional signals. In addition, the uracil-containing sequences on the amino termini encoded the cleavage site for TEV protease (Life Technologies, Inc.), and those on the carboxy terminal encoded consecutive TAA nonsense codons.

Unpurified PCR products (about 30 ng) were mixed with the gel purified, linear rf1, rf2, or rf3 cloning vectors (about 50 ng) in a 10 μl reaction containing 1x REact 4 buffer (LTI) and 1 unit UDG (LTI). After 30 minutes at 37° C., 1 μl aliquots of each reaction were transformed into competent *E. coli* DH5α cells (LTI) and plated on agar containing 50 μg/ml kanamycin. Colonies were picked and analysis of miniprep DNA showed that the CAT gene had been cloned in reading frame 1 (pEZC3601)(FIG. 9C), reading frame 2 (pEZC3609)(FIG. 8D) and reading frame 3 (pEZC3617) (FIG. 8E), and that the phoA gene had been cloned in reading frame 1 (pEZC3606)(FIG. 8F), reading frame 2 (pEZC3613)(FIG. 8G) and reading frame 3 (pEZC3621) (FIG. 8H).

Part IV: Subhloning of CAT or phoA from UDG Cloning Vectors into a GST Fusion Vector

Plasmids encoding fusions between GST and either CAT or phoA in all three reading frames were constructed by recombinational cloning as follows. Miniprep DNA of GST vector donor pEZC3501(FIG. 8B) (derived from Pharmacia plasmid pGEX-2TK as described above) was linearized with Cla I. About 5 ng of vector donor were mixed with about 10 ng each of the appropriate circular gene donor vectors containing CAT or phoA in 8 μl reactions containing buffer and recombination proteins Int, Xis, and IHF (above). After incubation, 1 μl of each reaction was transformed into *E. coli* strain DH5α and plated on ampicillin, as presented in Table 7.

TABLE 7

DNA	Colonies (10% of each transformation)
Linear vector donor (pEZC3501/Cla)	0
Vector donor + CAT rf1	110
Vector donor + CAT rf2	71
Vector donor + CAT rf3	148
Vector donor + phoA rf1	121
Vector donor + phoA rf2	128
Vector donor + phoA rf3	31

Part V: Expression of Fusion Proteins

Two colonies from each transformation were picked into 2 ml of rich medium (CircleGrow, Bio101 Inc.) in 17×100 mm plastic tubes (Falcon 2059, Becton Dickinson) containing 100 μg/ml ampicillin and shaken vigorously for about 4 hours at 37° C., at which time the cultures were visibly turbid. One ml of each culture was transferred to a new tube containing 10 μl of 10% (w/v) IPTG to induce expression of GST. After 2 hours additional incubation, all cultures had about the same turbidity; the A600 of one culture was 1.5. Cells from 0.35 ml each culture were harvested and treated with sample buffer (containing SDS and β-mercaptoethanol) and aliquots equivalent to about 0.15 A600 units of cells were applied to a Novex 4–20% gradient polyacrylamide gel. Following electrophoresis the gel was stained with Coomassie blue.

Results: Enhanced expression of single protein bands was seen for all 12 cultures. The observed sizes of these proteins correlated well with the sizes predicted for GST being fused (through attB recombination sites without stop codons) to CAT or phoA in three reading frames: CAT rf1=269 amino acids; CAT rf2=303 amino acids; CAT rf3=478 amino acids; phoA rf1=282 amino acids; phoA rf2=280 amino acids; and phoA rf3=705 amino acids.

Analysis: Both CAT and phoA genes were subcloned into a GST fusion vector in all three reading frames, and expression of the six fusion proteins was demonstrated.

While the foregoing invention has been described in some detail for purposes of clarity and understanding, it will be appreciated by one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the invention and appended claims. All patents and publications cited herein are entirely incorporated herein by reference.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i i i) NUMBER OF SEQUENCES: 35

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:1:

RKYCWGCTTT YKTRTACNAA STSGB 25

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:2:

AGCCWGCTTT YKTRTACNAA CTSGB 25

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:3:

-continued

GTTCAGCTTT CKTRTACNAA CTSGB 25

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:4:

AGCCWGCTTT CKTRTACNAA GTSGB 25

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GTTCAGCTTT YKTRTACNAA GTSGB 25

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AGCCTGCTTT TTTGTACAAA CTTGT 25

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:7:

AGCCTGCTTT CTTGTACAAA CTTGT 25

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:8:

ACCCAGCTTT CTTGTACAAA CTTGT 25

(2) INFORMATION FOR SEQ ID NO:9:

-continued

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GTTCAGCTTT TTTGTACAAA CTTGT 25

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GTTCAGCTTT CTTGTACAAA CTTGT 25

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GTTCAGCTTT CTTGTACAAA GTTGG 25

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:12:

AGCCTGCTTT TTTGTACAAA GTTGG 25

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:13:

AGCCTGCTTT CTTGTACAAA GTTGG 25

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

-continued

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:14:

ACCCAGCTTT CTTGTACAAA GTTGG 25

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GTTCAGCTTT TTTGTACAAA GTTGG 25

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GTTCAGCTTT CTTGTACAAA GTTGG 25

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CCACCACAAA CGCGTCCATG GAATTACACT TTAATTTAG 39

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CCACCACAAG TCGACGCATG CCGACAGCCT TCCAAATGT 39

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 46 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:19:

-continued

GGCCGATTAC GATATCCCAA CGACCGAAAA CCTGTATTTT CAGGGT 46

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CAGGTTTTCG GTCGTTGGGA TATCGTAATC 30

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 47 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GGCCAGATTA CGATATCCCA ACGACCGAAA ACCTGTATTT TCAGGGT 47

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CAGGTTTTCG GTCGTTGGGA TATCGTAATC T 31

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 48 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GGCCAAGATT ACGATATCCC AACGACCGAA AACCTGTATT TTCAGGGT 48

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CAGGTTTTCG GTCGTTGGGA TATCGTAATC TT 32

(2) INFORMATION FOR SEQ ID NO:25:

-continued

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:25:

ACCGTTTACG TGGAC 15

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:26:

TCGAGTCCAC GTAAACGGTT CCCACTTATT A 31

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:27:

UAUUUUCAGG GUATGGAGAA AAAAATCACT GGATATACC 39

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:28:

UCCCACUUAU UACGCCCCGC CCTGCCACTC ATC 33

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:29:

UAUUUUCAGG GUATGCCTGT TCTGGAAAAC CGG 33

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

-continued

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:30:

UCCCACUUAU UATTTCAGCC CCAGGGCGGC TTTC 34

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:31:

AGCCTGCTTT TTTATACTAA CTTGA 25

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:32:

TCAAGTTAGT ATAAAAAAGC AGGCT 25

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:33:

ACAAGTTTGT ACAAAAAAGC AGGCT 25

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:34:

ACAAGTTTGT ACAAGAAAGC AGGCT 25

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

-continued

(i i) MOLECULE TYPE: cDNA

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:35:

ACCACTTTGT ACAAGAAAGC TGGGT 25

What is claimed is:

1. A Vector Donor DNA molecule comprising a first DNA segment and a second DNA segment, said first or second DNA segment containing at least one Selectable marker, wherein

i) said first or second DNA segment is flanked by at least a first and a second recombination site; and

ii) said first recombination site and said second recombination site do not recombine with each other.

2. The Vector Donor DNA molecule according to claim **1**, wherein said Selectable marker comprises at least one inactive fragment of the Selectable marker, wherein the inactive fragment reconstitutes a functional Selectable marker when recombined across said first or second recombination site with a futher DNA segment comprising another inactive fragment of the Selectable marker.

3. The Vector Donor DNA molecule of claim **1**, wherein at least one of said recombination sites is derived from at least one recombination site selected from the group consisting of attB, attP, attL, and attR.

4. The Vector Donor DNA molecule according to claim **1**, wherein the Selectable marker comprises at least one DNA segment selected from the group consisting of:

(i) a DNA segment that encodes a product that provides resistance against otherwise toxic compounds;

(ii) a DNA segment that encodes a heteolosous product;

(iii) a DNA segment that encodes a product that suppresses the activity of a gene product;

(iv) a DNA segment that encodes a product that is identifiable;

(v) a DNA segment tat encodes a product that inhibits a cell function;

(vi) a DNA segment that inhibits the activity of any of the DNA segments of (i)–(v) above;

(vii) a DNA segment that binds a product that modifies a substrate;

(viii) a DNA segment that provides for the isolation of a desired molecule;

(ix) a DNA segment that encodes a specific nucleotide recognition sequence which is recognized by an enzyme; and

(x) a DNA segment that, when deleted, confers sensitivity to cell-killing by particular compounds.

5. The Vector Donor DNA molecule according to claim **4**, wherein said Selectable marker comprises at least one marker selected from the group consisting of an antibiotic resistance gene, a tRNA gene, an auxotrophic marker, a toxic gene, a phenotypic marker, an antisense oligonucleotide; a restriction endonuclease; a restriction endonuclease cleavage site, an enzyme cleavage site, a protein binding site; and a sequence complementary to a PCR primer sequence.

6. The Vector Donor DNA molecule according to claim **1**, wherein said recombination site comprises a DNA sequence selected from the group consisting of:

(a) RKYCWGCTTTYKTRTACNAASTSGB (m-att) (SEQ ID NO:1);

(b) AGCCWGCTTTYKTRTACNAACTSGB (m-attB) (SEQ ID NO:2);

(c) GTTCAGCTTTCKTRTACNAACTSGB (m-attR) (SEQ ID NO:3);

(d) AGCCWGCTTTCKTRTACNAAGTSGB (m-attL) (SEQ ID NO:4);

(e) GTTCAGCTTTYKTRTACNAAGTSGB(m-attP1) (SEQ ID NO:5);

and a corresponding or complementary DNA or RNA sequence, wherein R=A or G; K=G or T/U; Y=C or T/U; W=A or T/U; N=A, C, or G or T/U; S=C or G; and B=C, G or T/U.

7. The Vector Donor DNA molecule according to claim **6**, wherein said DNA sequence comprises a sequence selected from the group consisting of:

(a) AGCCTGCTTTTTTGTACAAACTTGT (attB1) (SEQ ID NO:6);

(b) AGCCTGCTTTCTTGTACAAACTTGT (attB2) (SEQ ID NO:7);

(c) ACCCAGCTTTCTTGTACAAACTTGT (attB3) (SEQ ID NO:8);

(d) GTTCAGCTTTTTTGTACAAACTTGT (attR1) (SEQ ID NO:9);

(e) GTTCAGCTTTCTTGTACAAACTTGT (attR2) (SEQ ID NO:10);

(f) GTTCAGCTTTCTTGTACAAAGTTGG (attR3) (SEQ ID NO:11);

(g) AGCCTGCTTTTTTGTACAAAGTTGG (attL1) (SEQ ID NO:12);

(h) AGCCTGCTTTCTTGTACAAAGTTGG (attL2) (SEQ ID NO:13);

(i) ACCCAGCTTTCTTGTACAAAGTTGG (attL3) (SEQ ID NO:14);

(j) GTTCAGCTTTTTTGTACAAAGTTGG(attP1) (SEQ ID NO:15);

(k) GTTCAGCTTTCTTGTACAAAGTTGG (attP2,P3) (SEQ ID NO:16);

and a corresponding or complementary DNA or RNA sequence.

8. An Insert Donor DNA molecule, comprising a first DNA segment flanked by at least a first recombination site and a second recombination site, wherein said first and second recombination sites do not recombine with each other.

9. The Insert Donor DNA molecule according to claim **8**, wherein said desired DNA segment codes for at least one marker selected from the group consisting of a cloning site, a restriction site, a promoter, an operon, an origin of replication, a functional DNA, an antisense RNA, a PCR fragment, a protein and a protein fragment.

10. The Insert Donor DNA molecule according to claim **8**, wherein said recombination site comprises a DNA sequence selected from the group consisting of:

(a) RKYCWGCTTTYKTRTACNAASTSGB (m-att) (SEQ ID NO:1);

(b) AGCCWGCTTTYKTRTACNAACTSGB (m-attB) (SEQ ID NO:2);

(c) GTTCAGCTTTCKTRTACNAACTSGB (m-attR) (SEQ ID NO:3);

(d) AGCCWGCTTTCKTRTACNAAGTSGB (m-attL) (SEQ ID NO:4);

(e) GTTCAGCTTTYKTRTACNAAGTSGB(m-attP1) (SEQ ID NO:5);

and a corresponding or complementary DNA or RNA sequence, wherein R=A or G; K=G or T/U; Y=C or T/U; W=A or T/U; N=A C, or G or T/U; S=C or G; and B=C, G or T/U.

11. The Insert Donor DNA molecule according to claim 10, wherein said DNA sequence comprises a sequence selected from the group consisting of:

(a) AGCCTGCTTTTTTGTACAAACTTGT (attB1) (SEQ ID NO:6);

(b) AGCCTGCTTTCTTGTACAAACTTGT (attB2) (SEQ ID NO:7);

(c) ACCCAGCTTTCTTGTACAAACTTGT (attB3) (SEQ ID NO:8);

(d) GTTCAGCTTTTTTGTACAAACTTGT (attR1) (SEQ ID NO:9);

(e) GTTCAGCTTTCTTGTACAAACTTGT (attR2) (SEQ ID NO:10);

(f) GTTCAGCTTTCTTGTACAAAGTTGG (attR3) (SEQ ID NO:11);

(g) AGCCTGCTTTTTTGTACAAAGTTGG (attL1) (SEQ ID NO:12);

(h) AGCCTGCTTTCTTGTACAAAGTTGG (attL2) (SEQ ID NO:13);

(i) ACCCAGCTTTCTTGTACAAAGTTGG (attL3) (SEQ ID NO:14);

(j) GTTCAGCTTTTTTGTACAAAGTTGG(attP1) (SEQ ID NO:15);

(k) GTTCAGCTTTCTTGTACAAAGTTGG (attP2,P3) (SEQ ID NO:16);

and a corresponding or complementary DNA or RNA sequence.

12. A kit comprising at least one Vector Donor DNA molecule comprising at least a first DNA segment and a second DNA segment, said first or second DNA segment containing at least one Selectable marker, wherein said first or second DNA segment is flanked by at least a first and a second recombination site, that do not recombine with each other.

13. The kit according to claim 12, further comprising at least one Insert Donor DNA molecule comprising a desired DNA segment flanked by at least a first recombination site and a second recombination site that do not recombine with each other.

14. The kit according to claim 12, wherein said recombination site comprises a DNA sequence selected from the group consisting of:

(a) RKYCWGCTTTYKTRTACNAASTSGB (m-att) (SEQ ID NO:1),

(b) AGCCWGCTTTYKTRTACNAACTSGB (m-attB) (SEQ ID NO:2);

(c) GTTCAGCTTTCKTRTACNAACTSGB (m-attR) (SEQ ID NO:3);

(d) AGCCWGCTTTCKTRTACNAAGTSGB (m-attL) (SEQ ID NO:4);

(e) GTTCAGCTTTYKTRTACNAAGTSGB(m-attP1) (SEQ ID NO:5);

and a corresponding or complementary DNA or RNA sequence, wherein R=A or G; K=G or T/U; Y=C or T/U; W=A or T/U; N=A, C, or G or T/U, S=C or G; and B=C, G or T/U.

15. The kit according to claim 14, wherein said DNA sequence comprises a sequence selected from the group consisting of:

(a) AGCCTGCTTTTTTGTACAAACTTGT (attB1) (SEQ ID NO:6);

(b) AGCCTGCTTTCTTGTACAAACTTGT (attB2) (SEQ ID NO:7);

(c) ACCCAGCTTTCTTGTACAAACTTGT (attB3) (SEQ ID NO:8);

(d) GTTCAGCTTTTTTGTACAAACTTGT (attR1) (SEQ ID NO:9);

(e) GTTCAGCTTTCTTGTACAAACTTGT (attR2) (SEQ ID NO:10);

(f) GTTCAGCTTTCTTGTACAAAGTTGG (attR3) (SEQ ID NO:11);

(g) AGCCTGCTTTTTTGTACAAAGTTGG (attL1) (SEQ ID NO:12);

(h) AGCCTGCTTTCTTGTACAAAGTTGG (attL2) (SEQ ID NO:13);

(i) ACCCAGCTTTCTTGTACAAAGTTGG (attL3) (SEQ ID NO:14);

(j) GTTCAGCTTTTTTGTACAAAGTTGG(attP1) (SEQ ID NO:15);

(k) GTTCAGCTTTCTTGTACAAAGTTGG (antP2,P3) (SEQ ID NO:16);

and a corresponding or complementary DNA or RNA sequence.

16. A recombinant nucleic acid molecule, comprising at least one DNA segment comprising at least a first and a second recombination site flanking a Selectable marker or at least one desired DNA segment, wherein at least one of said first and said second recombination sites comprises a core region that enhances recombination efficiency or specificity in vitro in the formation of a Cointegrate DNA or a Product DNA, and wherein said first and second sites do not recombine with each other.

17. A composition, comprising the recombinant nucleic acid molecule according to claim 16, and a carrier.

18. The recombinant nucleic acid molecule according to claim 16, wherein said recombination sites confer at least one enhancement selected from the group consisting of (i) enhancing excisive recombination; (ii) enhancing integrative recombination; (iii) decreasing the requirement for host factors; (iv) increasing the efficiency of the formation reaction by recombination of said Cointegrate DNA or of said Product DNA; (v) increasing the specificity of the formation reaction by recombination of said Cointegrate DNA or of said Product DNA; and (vi) increasing the specificity or yield of a subsequent recombination reaction of, or subsequent isolation of, the Product DNA.

19. The recombinant nucleic acid molecule according to claim 18, wherein said att site is at least one selected from the group consisting of att1, att2 and att3.

20. The recombinant nucleic acid molecule according to claim 16, wherein said at least one of said recombination sites is from at least one att recombination site.

21. The recombinant nucleic acid molecule according to claim 20, wherein the att site is at least one selected from the group consisting of attB, attP, attL and attR.

22. The recombinant nucleic acid according to claim 16, wherein said core region comprises a DNA sequence selected from the group consisting of:

(a) RKYCWGCTTTYKTRTACNAASTSGB(m-att) (SEQ ID NO:1);

(b) AGCCWGCTTTYKTRTACNAACTSGB (m-attB) (SEQ ID NO:2);

(c) GTTCAGCTTTCKTRTACNAACTSGB (m-attR) (SEQ ID NO:3);

(d) AGCCWGCTTTCKTRTACNAAGTSGB (m-attL) (SEQ ID NO:4);

(e) GTTCAGCTTTYKTRTACNAAGTSGB(m-attP1) (SEQ ID NO:5);

and a corresponding or complementary DNA or RNA sequence, wherein R=A or G; K=G or T/U; Y=C or T/U; W=A or T/U; N=A or C or G or T/U; S=Cor G; and B=C or G or T/U.

23. The recombinant nucleic acid according to claim **22**, wherein said core region comprises a DNA sequence selected from the group consisting of:

(a) AGCCTGCTTTTTTGTACAAACTTGT (attB1) (SEQ ID NO:6);

(b) AGCCTGCTTTCTTGTACAAACTTGT (attB2) (SEQ ID NO:7);

(c) ACCCAGCTTTCTTGTACAAACTTGT (attB3) (SEQ ID NO:8);

(d) GTTCAGCTTTTTTGTACAAACTTGT (attR1) (SEQ ID NO:9);

(e) GTTCAGCTTTCTTGTACAAACTTGT (attR2) (SEQ ID NO:10);

(f) GTTCAGCTTTCTTGTACAAAGTTGG (attR3) (SEQ ID NO:11);

(g) AGCCTGCTTTTTTGTACAAAGTTGG (attL1) (SEQ ID NO:12);

(h) AGCCTGCTTTCTTGTACAAAGTTGG (attL2) (SEQ ID NO:13);

(i) ACCCAGCTTTCTTGTACAAAGTTGG (attL3) (SEQ ID NO:14);

(j) GTTCAGCTTTTTTGTACAAAGTTGG(attP1) (SEQ ID NO:15);

(k) GTTCAGCTTTCTTGTACAAAGTTGG (attP2,P3) (SEQ ID NO:16);

and a corresponding or complementary DNA or RNA sequence.

24. A kit, comprising the recombinant nucleic acid according to claim **16**.

25. The kit according to claim **24**, further comprising at least one recombination protein that recognizes at least one of said recombination sites.

26. A recombinant nucleic acid molecule, comprising

at least one recombination site comprising at least one nucleic acid sequence having at least one of SEQ ID NOS:1–16, or a complementay DNA sequence or a corresponding RNA sequence.

27. The method according to claim **26**, wherein said nucleic acid sequence is selected from the group consisting of:

(a) RKYCWGCTTTYKTRTACNAASTSGB (m-att) (SEQ ID NO:1);

(b) AGCCWGCTTTYKTRTACNAACTSGB (m-attB) (SEQ ID NO:2);

(c) GTTCAGCTTTCKTRTACNAACTSGB (m-attR) (SEQ ID NO:3);

(d) AGCCWGCTTTCKTRTACNAAGTSGB (m-attL) (SEQ ID NO:4);

(e) GTTCAGCTTTYKTRTACNAAGTSGB(m-attP1) (SEQ ID NO:5);

and a corresponding or complementary DNA or RNA sequence, wherein R=A or G; K=G or T/U; Y=C or T/U; W=A or T/U; N=A, C, G or T/U; S=C or G; and B=C, G or T/U.

28. The method according to claim 27, wherein said nucleic acid sequence is selected from the group consisting of:

(a) AGCCTGCTTTTTTGTACAAACTTGT (attB1) (SEQ ID NO:6);

(b) AGCCTGCTTTCTTGTACAAACTTGT (attB2) (SEQ ID NO:7);

(c) ACCCAGCTTTCTTGTACAAACTTGT (attB3) (SEQ ID NO:8);

(d) GTTCAGCTTTTTTGTACAAACTTGT (attR1) (SEQ ID NO:9);

(e) GTTCAGCTTTCTTGTACAAACTTGT (attR2) (SEQ ID NO:10);

(f) GTTCAGCTTTCTTGTACAAAGTTGG (attR3) (SEQ ID NO:11);

(g) AGCCTGCTTTTTTGTACAAAGTTGG (attL1) (SEQ ID NO:12);

(h) AGCCTGCTTTCTTGTACAAAGTTGG (attL2) (SEQ ID NO:13);

(i) ACCCAGCTTTCTTGTACAAAGTTGG (attL3) (SEQ ID NO:14);

(j) GTTCAGCTTTTTTGTACAAAGTTGG(attP1) (SEQ ID NO:15);

(k) GTTCAGCTTTCTTGTACAAAGTTGG (attP2,P3) (SEQ ID NO:16);

and a corresponding or complementary DNA or RNA sequence.

29. A method of making a Cointegrate DNA molecule, comprising combining in vitro.

(i) an Insert Donor DNA molecule, comprising a desired DNA segment flanked by a first recombination site and a second recombination site, wherein the first and second recombination sites do not recombine with each other;

(ii) a Vector Donor DNA molecule containing a third recombination site and a fourth recombination site, wherein the third and fourth recombination sites do not recombine with each other; and

(iii) at least one site specific recombination protein capable of recombining said first and third recombinational sites said second and fourth recombinational sites;

thereby allowing recombination to occur, so as to produce a Cointegrate DNA molecule comprising said first and third or said second and fourth recombination sites.

30. The method as claimed in claim **29**, wherein the Vector Donor DNA molecule comprises a vector segment flanked by said third and said fourth recombination sites.

31. The method as claimed in claim 29, wherein the Vector Donor DNA molecule further comprises (a) a toxic gene and (b) a Selectable marker, wherein said toxic gene and said Selectable marker are on different DNA segments, the DNA segments being separated from each other by at least two recombination sites.

32. The method as claimed in claim **29**, wherein the Vector Donor DNA molecule further comprises (a) a repression cassette and (b) a Selectable marker that is repressed by the repressor encoded by said repression cassette, and wherein the Selectable marker and the repression cassette are on different DNA segments, the DNA segments being separated from each other by at least two recombination sites.

33. The method as claimed in claim **29**, wherein at least one of said Insert Donor DNA molecule and said Vector Donor DNA molecule is comprised of a circular DNA molecule.

34. The method as claimed in claim 29, wherein at least one of said Insert Donor DNA molecule and said Vector Donor DNA molecule is comprised of a linear DNA molecule.

35. The method of claim 29, further comprising the step of

producing a Product DNA molecule from said Cointegrate DNA by recombining at least one of (i) said first and third, or (ii) said second and fourth, recombination sites, said Product DNA comprising said desired DNA segment.

36. The method according to claim 35, wherein said method also produces a Byproduct DNA molecule, wherein said Byproduct DNA molecule does not contain said desired DNA segment and is produced with said Product DNA.

37. The method according to claim 35, further comprising the step of selecting the Product DNA molecule.

38. A recombinant nucleic acid molecule comprising at least a first and a second recombination site flanking at least one DNA segment containing at least one Selectable marker, wherein said first and second recombination sites do not recombine with each other.

39. The recombinant nucleic acid molecule of claim 38, wherein said selectable marker is selected from the group consisting of:

(i) a DNA segment that encodes a product that provides resistance against otherwise toxic compounds;

(ii) a DNA segment that encodes a heterologous product;

(iii) a DNA segment that encodes a product that suppresses the activity of a gene product;

(iv) a DNA segment that encodes a product that is identifiable;

(v) a DNA segment that encodes a product that inhibits a cell function;

(vi) a DNA segment that inhibits the activity of any of the DNA segments of (i) to (v) above;

(vii) a DNA segment that binds a product that modifies a substrate;

(viii) a DNA segment that provides for isolation of a desired molecule;

(ix) a DNA segment that encodes a specific nucleotide recognition sequence which is recognized by an enzyme; and

(x) a DNA segment that, when deleted, confers sensitivity to cell killing by a particular compound.

40. The recombinant nucleic acid molecule of claim 38, wherein said selectable marker is selected from the group consisting of an antibiotic resistance gene, a tRNA gene, an auxotrophic marker, a toxic gene, a phenotypic marker, an antisense oligonucleotide, a restriction endonuclease, a restriction endonuclease cleavage site, an enzyme cleavage site, a protein binding site, and a sequence complimentary to a PCR primer sequence.

41. A kit comprising the recombinant nucleic acid molecule of claim 38.

42. The recombinant nucleic acid according to claim 38, wherein said DNA segment comprises a cloning site.

43. The recombinant nucleic acid according to claim 42, wherein said nucleic acid contains at least one restriction enzyme site at said cloning site.

44. The recombinant nucleic acid according to claim 38, wherein said DNA segment further comprises an insert DNA molecule.

45. The recombinant nucleic acid according to claim 44, wherein said Insert DNA molecule codes for at least one marker selected from the group consisting of a restriction site, a promoter, an operon, an origin of replication, a functional DNA, an antisense RNA, a PCR fragment, a protein or a protein fragment.

46. The molecule according to claim 38, wherein said recombination site comprises a DNA sequence selected from the group consisting of:

(a) RKYCWGCTTTYKTRTACNAASTSGB (m-att) (SEQ ID NO:1);

(b) AGCCWGCTTTYKTRTACNAACTSGB (m-attB) (SEQ ID NO:2);

(c) GTTCAGCTTTCKTRTACNAACTSGB (m-attR) (SEQ ID NO:3);

(d) AGCCWGCTTTCKTRTACNAAGTSGB (m-attL) (SEQ ID NO:4);

(e) GTTCAGCTTTYKTRTACNAAGTSGB(m-attP1) (SEQ ID NO:5);

and a corresponding or complementary DNA or RNA sequence, wherein R=A or G; K=G or T/U; Y=C or T/U; W=A or T/U; N=A, C, G or T/U; S=C or G; and B=C, G or T/U.

47. The molecule according to claim 46, wherein said DNA sequence comprises a sequence selected from the group consisting of:

(a) AGCCTGCTTTTTTGTACAAACTTGT (attB1) (SEQ ID NO:6);

(b) AGCCTGCTTTCTTGTACAAACTTGT (attB2) (SEQ ID NO:7);

(c) ACCCAGCTTTCTTGTACAAACTTGT (attB3) (SEQ ID NO:8);

(d) GTTCAGCTTTTTTGTACAAACTTGT (attR1) (SEQ ID NO:9);

(e) GTTCAGCTTTCTTGTACAAACTTGT (attR2) (SEQ ID NO:10);

(f) GTTCAGCTTTCTTGTACAAAGTTGG (attR3) (SEQ ID NO:11);

(g) AGCCTGCTTTTTTGTACAAAGTTGG (attL1) (SEQ ID NO:12);

(h) AGCCTGCTTTCTTGTACAAAGTTGG (attL2) (SEQ ID NO:13);

(i) ACCCAGCTTTCTTGTACAAAGTTGG (attL3) (SEQ ID NO:14);

(j) GTTCAGCTTTTTTGTACAAAGTTGG(attP1) (SEQ ID NO:15);

(k) GTTCAGCTTTCTTGTACAAAGTTGG (attP2,P3) (SEQ ID NO:16);

and a corresponding or complementary DNA or RNA sequence.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 5,888,732
DATED : March 30, 1999
INVENTOR(S) : Hartley et al.

It is certified that error appears in the above-identified patent and that said Letters Patent is

Column 46,
Line 19, please delete "claim 6" and insert therefor -- claim 1 --.
Line 20, please delete "DNA sequence" and insert therefor -- recombination site --.

Column 47,
Line 10, please delete "claim 10" and insert therefor -- claim 8 --.
Line 10, please delete "DNA sequence" and insert therefor -- recombination site --.

Column 48,
Line 1, please delete "claim 14" and insert therefor -- claim 12 --.
Lines 1-2, please delete "DNA sequence" and insert therefor -- recombination site --.
Line 24, please delete "antP2,P3" and insert therefor -- attP2,P3 --.
Line 49, please delete "yield" and insert therefor -- yield --.
Line 55, after "wherein" and before "at", please delete "said".
Line 58, after "site is" please delete "at least one".

Column 49,
Line 11, please delete "claim 22" and insert therefor -- claim 16 --.
Line 48, please delete "complementay" and insert therefor -- complementary --.
Line 50, please delete "method" and insert therefor -- recombinant nucleic acid molecule --.

Column 50,
Line 1, please delete "method" and insert therefor -- recombinant nucleic acid molecule --.
Line 1, please delete "claim 27" and insert therefor -- claim 26 --.
Line 29, please delete "vitro." and insert therefor -- vitro: --.
Line 41, after "sites" and before "said", please insert -- and/or --.

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 5,888,732
DATED : March 30, 1999
INVENTOR(S) : Hartley et al.

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 52,
Line 30, please delete "claim 46" and insert therefor -- claim 38 --.
Line 31, please delete "DNA sequence" and insert therefor -- recombinant site --.

Signed and Sealed this

Seventh Day of August, 2001

Attest:

Nicholas P. Godici

Attesting Officer

NICHOLAS P. GODICI
Acting Director of the United States Patent and Trademark Office

EXHIBIT 5

THE JOURNAL OF BIOLOGICAL CHEMISTRY

Vol. 259, No. 20, Issue of October 25, pp. 12724-12732, 1984
Printed in U.S.A.

Purification and Properties of Int-h, a Variant Protein Involved in Site-specific Recombination of Bacteriophage λ*

(Received for publication, March 13, 1984)

Brenda J. Lange-Gustafson‡ and Howard A. Nash§

From the Laboratory of Neurochemistry, National Institute of Mental Health, Bethesda, Maryland 20205

Under physiological conditions, integration of λ DNA into the *Escherichia coli* chromosome requires the direct participation of only two proteins, the viral *int* gene product and *E. coli* integration host factor (IHF). A variant of the *int* gene has been isolated that permits integrative recombination in cells mutant for one of the two subunits of IHF (Miller, H. I., Mozola, M. A., and Friedman, D. I. (1980) *Cell* 20, 721–729). In the present work, we have purified Int-h, the product of this variant gene. In contrast to the wild-type *int* gene product (Int⁺), which produces almost no recombinants in the absence of IHF, purified Int-h protein sponsors reduced but significant levels of integrative recombination in the absence of any *E. coli* supplement. This shows that the *int* gene encodes all the information necessary for the elementary steps in recombination and implies that IHF functions as an accessory protein.

When supplemented by IHF, recombination promoted by Int-h resembles that promoted by Int⁺ in kinetics, stoichiometry of Int and IHF, and nature of the recombinant product. Under these conditions, Int-h uses supercoiled DNA more effectively than nonsupercoiled DNA as a substrate for recombination, as does Int⁺. However, in the absence of IHF, Int-h recombines supercoiled and nonsupercoiled substrates identically, indicating that IHF is an important part of the mechanism that senses the supercoiled state of the substrate DNA during recombination. A surprising difference in recombination carried out by Int-h in the presence or absence of IHF concerns the degree to which sites on the same circle recombine with one another as opposed to sites on sister molelcules. In the presence of IHF, Int-h favors intramolecular recombination, as does Int⁺. However, in the absence of IHF, Int-h almost exclusively promotes intermolecular recombination.

Bacteriophage λ has a specialized system for the integration of viral DNA into the bacterial chromosome. This system carries out a reciprocal recombination between a specific viral site, *attP*, and a specific host site, *attB*. The sequences and functional extents of these sites are known (for a recent review, see Ref. 1). A combination of genetic and biochemical experiments has shown that integrative recombination is carried out by two proteins: Int, the product of the viral *int* gene and IHF,[1] a protein that is composed of two polypeptides, the products of the *E. coli himA* and *hip* genes (1). Several studies have led to the conclusion that Int is the protein that carries out the breakage and rejoining steps in recombination. First, Int binds to the core region of *attP* and *attB*, the 15-base pair region of homology wherein the recombination crossover occurs (2). Moreover, Int has a topoisomerase activity that can relax supercoiled DNA (3) and can break attachment site DNA, albeit at low frequency, precisely at the nucleotides within the core that are involved in the crossover (4). Finally, Int can promote the exchange of a pair of strands in DNA assemblies that have been constructed to resemble recombination intermediates (5). Although these findings suggest that the role of Int is simply to promote strand exchange, other data suggest that it has additional roles. Chemical modification of Int can destroy recombination activity while leaving binding to the core and relaxing activity unchanged (4, 6). In addition, Int binds to portions of *attP* that are exterior to the core; this binding to the so-called arms of *attP* appears to be essential for recombination activity (2, 7). Analysis of the sequences protected by Int in the core and arm regions of *attP* indicates that Int is a bifunctional protein that recognizes two distinct binding sequences (6, 8).

The study of mutant proteins may be useful in dissecting the various ways in which Int protein promotes integrative recombination. In this paper, we begin the analysis of one variant, Int-h. This variant was isolated after selection for λ bacteriophage that could undergo site-specific recombination in an *E. coli* host that was mutant for IHF (9). The mutation proved to map in the *int* gene and *in vivo* studies indicated that the *int-h* allele produced a protein with an enhanced recombination potential. For example, in a strain deleted for *attB*, *int-h* was superior to *int*⁺ in promoting the integration of λ into secondary bacterial sites. In addition, *int-h* showed altered recombination potential for excision, the removal of integrated viral DNA (9). *In vivo* studies have not revealed the basis of the enhanced recombination efficiency of the *int-h* allele. It might be that the Int-h protein is altered in its capacity to interact with IHF, its affinity for core or arm binding sequences, its tendency to form nucleosome-like structures at attachment sites (10, 11), its intrinsic topoisomerase activity, etc. Since variation in any of these activities could provide a valuable probe for the analysis of the detailed mechanism of recombination, we have undertaken the study of the Int-h protein. This report presents our data on the cloning and purification of Int-h and our initial results on the characterization of the recombination capacity of this protein.

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Present address, Office of Extramural Project Review, National Institute of Mental Health, Rockville, MD 20857.

§ Present address, Laboratory of Molecular Biology, National Institute of Mental Health, Building 36/Room 1B08, Bethesda, MD 20205.

[1] The abbreviations used are: IHF, integration host factor; SDS, sodium dodecyl sulfate; TEMED, *N,N,N′,N′*-tetramethylethylenediamine; kb, kilobase; bp, base pair.

EXPERIMENTAL PROCEDURES

Materials

Bacteria and Bacteriophage—The hacteria used in this work were derivatives of the *E. coli* K12 strain N99. Strain HN356 (constructed by R. A. Weisberg, National Institutes of Health, Bethesda, MD) contains the *rec*B21 mutation. Strains K5185 and K572 (constructed by H. Miller, Genentech, Inc., San Francisco, CA) respectively contain a partial deletion, *him*A82, and a point mutation, *him*A42, in the gene for the α subunit of IHF. Strain K5248 (constructed by H. Miller) contains a point mutation, *hip*157, in the gene for the β subunit of IHF. Strain JD12 (constructed by K. Ahremski, E. I. duPont deNemours & Co., Wilmington, DE) contains both the *him*A42 and *hip*157 mutations.

Bacteriophage strain Y619 (constructed by H. Miller) is λ *h int-h intC*226 c1857; it was grown in strain K5185 and individual isolates were tested for the absence of *int*-promoted deletions by scoring sensitivity to EDTA (12). Bacteriophage strain G903 (constructed by S. Adhya, National Institutes of Health, Bethesda, MD) is λ *attB-attP int*2 *xis*1 *red*114 *imm*434 cI1 cIII28 cIIG11.

Plasmids—Plasmid pRSP2124 (13) is a derivative of colE1 that contains the TnA transposon; it was obtained from L. Enquist, E. I. duPont deNemours & Co., Wilmington, DE. Plasmid pC22642 (14) is pRSP2124 containing a *Eco*RI insert from λ *int* C226; pHN16 (this work) is the identical construct except that the *Eco*RI insert is from Y619. Recombination substrates were grown, labeled with [^{3}H]thymidine and purified as described (15). Plasmid pPA1, pBB105, and pBP1 are described in Ref. 15; plasmid pBP86 and pBP90 are described in Ref. 16. A detailed restriction map of the *attP* insert that is common to pPA1, pBP86, and pBP90 can be found in Ref. 11.

Proteins—Wild-type Int protein (Int$^+$) was purified as described (17) from strain HN695, a derivative of strain K5185 containing the plasmid pC22642. Int-h protein was purified from strain HN700, a derivative of strain K5185 containing the plasmid pHN16. Wild-type IHF was purified through Fraction V as described (15) from strain HN356. Crude extracts of wild-type and mutant IHF were made by sonication of strains N99, K5185, K572, K5248, and JD12 as described (17). Restriction endonucleases were purchased from Bethesda Research Laboratories and New England Biolabs.

Recombination Assays

Typical recombination mixtures (27 μl) contained 37 mM Tris-HCl (pH 7.4), 5.6 mM spermidine, 1.1 mM EDTA, 1.1 mg/ml bovine serum albumin, 25 to 75 mM KCl, purified IHF, and purified Int as indicated. In some reactions, purified IHF was replaced by 0.2 μl of sonic extract. The reaction mixtures also contained either 0.30 μg of a plasmid containing both *attP* and *attB* or 0.6 μg of an equimolar mixture of two plasmids, one containing *attP* and the other containing *attB*. Unless otherwise noted, the reactions were incubated for 1 h at 25 °C and then stopped as described for the intramolecular recombination assay in Ref. 15. Restriction of the recombined DNA was carried out with 10–50 units of endonuclease for 1 h at 37 °C. The samples were prepared for gel electrophoresis by addition of 5 μl of a solution containing 25% (w/v) Ficoll, 2% (w/v) SDS, and 0.1% (w/v) bromphenol blue and extracted once with about 100 μl of a 24:1 (v/v) mixture of chloroform and isoamyl alcohol. Agarose gel electrophoresis was carried out as described (15). To quantitate recombination, bands were visualized by ethidium bromide fluorescence, cut from the gel, solubilized at 90 °C with 1 ml of 5 M sodium perchlorate, and counted with 15 ml of Aquasol (New England Nuclear). In some experiments, recombination was assessed by transfer of fragments to nitrocellulose paper by the method of Southern and hybridized with ^{32}P-labeled DNA as described (11). Acrylamide gel electrophoresis and quantitation of the resulting DNA bands were carried out as described (11).

Other Methods

Topoisomerase Activity—Relaxation assays (21 μl) contained 62 mM Tris-HCl (pH 7.5), 67 mM KCl, 5.25 mM EDTA, 0.30 mg/ml bovine serum albumin, 1 μg of pPA1 plasmid DNA, and 1 unit of Fraction III Int as indicated. The reaction mixtures were incubated at 25 °C, stopped by addition of 2 μl of 10% (w/v) SDS, and diluted to 6.16 ml with a solution of 50 mM Tris-HCl (pH 8.0) containing 25 mM EDTA. To this was added 6.05 g of cesium chloride and 0.35 ml of ethidium bromide (10 mg/ml) and 2.5 μl of ^{14}C-labeled supercoiled pPA1 plasmid DNA. The mixture was centrifuged at 15 °C in a Beckman type 65 rotor for 60 h at 35,000 rpm. The gradient was fractionated from the bottom and counted. To assess the extent of relaxation, the separation between the peak of the ^{14}C-labeled marker DNA and the peak of the treated DNA was divided by the separation between the marker DNA and a sample of pPA1 plasmid DNA that had been completely relaxed by treatment with HeLa cell topoisomerase I (a gift of Dr. L. Liu, Johns Hopkins University, Baltimore, MD).

SDS Gel Electrophoresis—The gels were slabs 16 cm × 17 cm × 1.5 mm. The separating gel contained 18% (w/v) acrylamide, 0.5% bisacrylamide, 1 M urea, 375 mM Tris-HCl (pH 8.8), 2 mM EDTA, 0.1% SDS, 0.1% ammonium persulfate, and 0.66% (v/v) TEMED. The stacking gel contained 5% acrylamide, 0.13% bisacrylamide, 1 M urea, 125 mM Tris-HCl (pH 6.8), 2 mM EDTA, 0.1% SDS, 0.1% ammonium persulfate, and 0.08% TEMED. The running buffer was 0.19 M glycine, 1 M urea, 0.025 M Tris base, and 0.1% SDS. The samples were precipitated with trichloroacetic acid and washed with acetone as described (15). They were resuspended in 28 μl of 120 mM Tris-HCl (pH 6.8), 2.4% SDS, 4.8 M EDTA, 24% (v/v) glycerol, 0.007% bromphenol blue, and 2.25% β-mercaptoethanol. The samples were then heated to 90 °C for 2 min and electrophoresed at 30 mA for 4 h. The gel was stained with 0.05% Coomassie Brilliant Blue in a solution of 50% methanol and 7.5% glacial acetic acid for 2 h and then destained in 5% methanol containing 7.5% glacial acetic acid.

RESULTS

Enzyme Purification—In order to provide a rich source of Int-h protein, we cloned the *int-h* gene in a multicopy plasmid. As before (18), we employed a λ variant, *int*C226, in which the *int* gene is expressed constitutively from an altered phage promoter (19). The structure of the hybrid plasmid is shown in Fig. 1. Although cells containing this plasmid grow well, subcultures of the original isolate occasionally contain smaller plasmids. These probably represent deletions of the plasmid shown in Fig. 1 that are created by Int-h promoted recombination between *attP* and sequences on the plasmid that resemble *attB*. To minimize the formation of deletions, we transferred the plasmid from its original host, a wild-type *E. coli*, to K5185, a strain carrying a deletion in the *him*A gene. Site-specific recombination is substantially reduced in this strain (see below) and, accordingly, we find the hybrid plasmid is considerably more stable. For the experiments reported in this paper, Int-h was purified from K5185 containing the hybrid plasmid.

To assay Int-h activity, we measured integrative recombination *in vitro* by *Eco*RI restriction of pBP86, a plasmid substrate that contains both *attP* and *attB*. Two kinds of reaction mixtures were used. In the standard assay, a source of Int-h was supplemented with a crude extract from wild-type *E. coli*. Using this assay, the purification of Int-h activity proceeded exactly as described for wild-type Int protein (17, 18). The yield and specific activity of the highly purified Int-h (Table I) are not substantially different from those found after purification of Int$^+$ protein (17). In a second assay, a source of Int-h was supplemented with a crude extract from *E. coli* carrying a point mutation in the *him*A gene. This assay measures a specific quality of the Int-h protein, *i.e.* the capac-

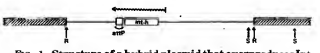


FIG. 1. **Structure of a hybrid plasmid that overproduces Int-h.** The plasmid, pHN16, consists of the cloning vector pRSP2124 and an *Eco*RI fragment containing the *int-h* allele; only the λ insert and flanking vector sequences are shown. The position of the *int* gene, the phage attachment site *attP*, and target sites for *Eco*RI (*R*) and *Sma*I (*S*) endonucleases are shown. The *left-pointing arrow* indicates the transcript expected to govern expression of *int* gene; it begins at the start point of the P_{int} promoter (made constitutive by the *int*-C226 mutation) and ends at the t_{int} terminator (31, 32). ▤, vector sequences.

TABLE I

Purification of Int-h protein

	Volume	Protein	Int-h	Specific activity
	ml	*mg*	*units*[a] $\times 10^{-2}$	*units/mg*
I. Crude extract[b]	75	2,325	2,250	967
II. Differential salt precipitation	74	—[c]	1,110	—
III. Phosphocellulose	12	12	990	82,500
IV. Calcium-phosphate cellulose	8	6	190	31,666

[a] The minimum amount of Int that produces maximal recombination (17).

[b] The yield from 75 liters of culture grown to midlog phase.

[c] —, not determined.

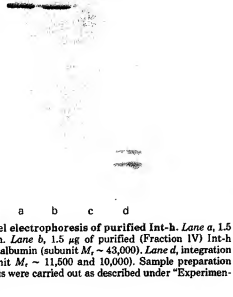


FIG. 2. SDS gel electrophoresis of purified Int-h. *Lane a*, 1.5 μg of Int$^+$ protein. *Lane b*, 1.5 μg of purified (Fraction IV) Int-h protein. *Lane c*, ovalbumin (subunit $M_r \sim 43{,}000$). *Lane d*, integration host factor (subunit $M_r \sim 11{,}500$ and 10,000). Sample preparation and electrophoresis were carried out as described under "Experimental Procedures."

ity to carry out recombination in the presence of mutant IHF. Wild-type Int protein shows almost no activity in this assay, whereas crude extracts containing Int-h produce readily detectable levels of recombination. During the purification of Int-h, the ratio of activities in the two assays remained constant (data not shown). Unless otherwise stated, all results reported in this paper are from experiments that use Int-h or Int$^+$ purified through Step IV of Table I and Ref. 17.

The purified protein is quite stable. We routinely add bovine serum albumin (2 mg/ml) to our purified proteins; under these conditions, Int-h activity is stable for at least 1 year at −70 °C. As found for purified wild-type Int protein, fractions containing Int-h protein without added bovine serum albumin show diminished activity after repeated freezing and thawing. The purified Int-h protein is nearly homogeneous. As shown in Fig. 2, SDS gel electrophoresis of the purified material shows a prominent major band that comigrates with purified Int$^+$ protein at an apparent $M_r \sim 40{,}000$. On careful examination, some minor bands are evident; these are similar in molecular weight and intensity to those seen in preparations of wild-type Int purified from a K5185 derivative. The identity in size as well as the similarity in purification of Int-h and Int$^+$ indicate that the *int-h* mutation does not radically alter the Int polypeptide. This hypothesis is supported by the similar sensitivity of the two purified proteins to inactivation. Both are readily inactivated either by incubation at 45 °C or by exposure to *N*-ethylmaleimide (data not shown).

TABLE II

Recombination in vitro and in vivo with Int-h and Int$^+$

In vitro recombination with a plasmid substrate pBP86 was carried out with purified Int-h or Int$^+$ supplemented by sonicates of the indicated *E. coli* strains. Reaction conditions, EcoRI restriction, and Southern blotting analysis were carried out as described under "Experimental Procedures." Recombination was quantitated by comparison of the intensity of the 8.1-kb recombinant bands (see Fig. 4 for a detailed restriction map); serial dilutions of each reaction mixture were analyzed to facilitate comparison. *In vivo* recombination was determined as the fraction of recombinant progeny following infection of cells containing either pHN16 or pC22642 with λ *attB-attP*, strain G903. The protocol for growth and analysis of this phage is described in Ref. 12.

Source of IHF	Relative recombination[a]			
	In vitro		*In vivo*	
	Int-h	Int$^+$	Int-h	Int$^+$
Wild type	1.0	1 (50%)	1.33	1 (60%)
*him*A42	0.33	0.004	0.66	0.057
*him*A82	0.10	<0.001	0.13	0.003[b]
None	0.10	0.002	NA[c]	NA

[a] The recombination observed with wild-type Int and IHF is assigned a value of 1.0; the actual conversion of substrate to recombinant under these conditions is given in parentheses.

[b] This value is at least 10-fold higher than that observed for a control infection that lacked a source of Int.

[c] NA, not applicable.

Recombination Promoted by Int-h—Table II compares the efficiency of recombination promoted by Int-h or Int$^+$ in the presence of different sources of crude IHF. As noted above, Int-h promotes efficient recombination when supplemented either with an extract of cells carrying the point mutation *him*A42 or supplemented with an extract of wild-type cells. By contrast, Int$^+$ yields very little recombination with the mutant extract (line 1 *versus* 2). Int-h cannot utilize all mutant extracts equally well. Extracts from cells carrying a deletion mutation, *him*A82, assist Int-h promoted recombination less than one-third as well as do extracts from *him*A42 (Table II, line 2 *versus* 3). In addition, extracts from cells carrying the point mutation *hip*157 or the double mutation *him*A42 *hip*157 are similar to extracts from *him*A82 cells in their capacity to assist Int-h promoted recombination (data not shown). Compared to these extracts, the enhanced recombination seen when Int-h is supplemented with extracts from *him*A42 cells suggests that the *him*A42 mutation has not completely inactivated the α subunit of IHF and that Int-h is better able than Int$^+$ to utilize the residual activity. The relative capacities of Int-h and Int$^+$ to promote recombination with various sources of IHF are not changed by altering the amount of Int protein, the amount of crude IHF, or the time of incubation (data not shown).

The recombination observed with extracts from cells containing the *him*A82 mutation, a deletion of the *him*A gene, shows that Int-h can promote recombination in the complete absence of a functional *him*A gene product. This result was unexpected because it had been reported earlier (9) that, *in vivo*, the *int-h* allele could not suppress the recombination defect of a strain bearing a deletion of the *him*A gene.[2] We

[2] Recent experiments show that in *him*A deletion strains, *int-h* can promote a low level of excision of λ from a secondary bacterial site (R. Weisberg; personal communication; D. Friedman; personal communication).

think this failure reflects the limited amount of Int protein made under the previous conditions since a small but readily detected amount of integrative recombination is observed *in vivo* after infection of *him*A82 cells when Int-h is provided from our overproducing plasmid (Table II).

The ability of Int-h to promote recombination with a wide variety of mutant IHF extracts suggests that this protein might have recombination activity in the total absence of IHF. This is confirmed in the last entry of Table II; this result indicates that, to the extent that our preparation is pure, Int-h can promote recombination by itself. The amount of IHF-independent recombination is not large, about 10% that seen when Int-h is supplemented with wild-type IHF. Note that this amount is similar to that seen when IHF is supplemented with crude extracts from a *him*A deletion or *hip* mutant strain. This means that other proteins found in *E. coli*, including either of the remaining wild-type subunits of IHF, cannot assist Int-h in promoting recombination. The capacity of Int-h to promote recombination by itself is also demonstrated in Fig. 3*A*. In the absence of IHF, Int-h (*lane a*) but not Int$^+$ (*lane f*) produces detectable recombinants. It should be pointed out that the mobility of the recombinant fragment produced in the absence of IHF is identical to that produced in its presence (*cf. lanes a* and *b*). This implies that the breakage and reunion caused by Int-h acting in the absence of IHF occurs at the same sites as observed in the standard recombination reaction.

The remainder of Fig. 3*A* presents the response of Int$^+$ and Int-h to increasing amounts of IHF that has been purified from wild-type *E. coli*. Both Int proteins are stimulated by IHF. We estimate that the amount of IHF needed to maximally stimulate Int$^+$ is about 3-fold greater than that required to stimulate Int-h. This modest difference indicates that Int-h is not greatly altered in its affinity for or capacity to utilize IHF. Fig. 3*B* shows the effect of adding increasing amounts of Int$^+$ or Int-h to a fixed, saturating amount of IHF. Similar amounts of the two proteins produce similar levels of recombination. This means that the Int-h mutation has not altered the number of Int molecules required to carry out recombination. No less than 35 Int-h molecules are needed per recombination event. However, as in our earlier studies with Int$^+$ (15), we do not know the extent to which this stoichiometry reflects inactive protein in our purified preparation.

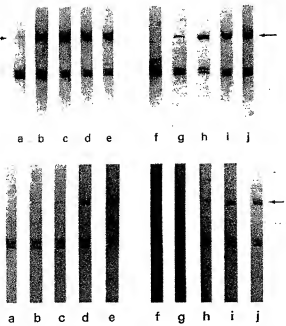


FIG. 3. **Recombination promoted by varying amounts of Int and IHF.** Supercoiled pBP86 substrate DNA was incubated in the presence of 50 mM KCl for 25 min as described under "Experimental Procedures" with proteins as indicated. In *panel A*, the reactions contained 6.0 μg/ml of Int-h (*a–e*) or Int$^+$ (*f–j*) and different concentrations of IHF: *a* and *f*, 0.0 μg/ml; *b* and *g*, 0.125 μg/ml; *c* and *h*, 0.5 μg/ml; *d* and *i*, 1.25 μg/ml; *e* and *j*, 2.5 μg/ml. In *panel B*, IHF was fixed at 2.5 μg/ml and either Int-h (*a–e*) or Int$^+$ (*f–j*) was varied: *a* and *f*, 0.15 μg/ml; *b* and *g*, 0.30 μg/ml; *c* and *h*, 0.60 μg/ml; *d* and *i*, 1.20 μg; *e* and *j*, 6.00 μg/ml. A *pair* of *arrows* indicates the position of the 8.1-kb recombinant fragment.

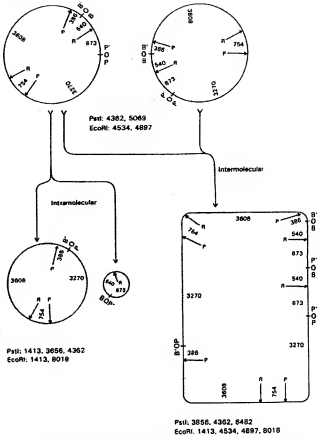


FIG. 4. **Restriction maps for analysis of intermolecular and intramolecular recombination.** At the *top* are shown two identical pBP86 substrate DNA circles. Attachment sites are written as *POP'* (*attP*) and *BOB'* (*attB*) where *O* represents the 15-bp core wherein the recombination crossover takes place. The position of *Eco*RI (*R*) and *Pst*I (*P*) restriction sites are marked with *arrows*. Distances (in base pairs) between adjacent restriction and/or attachment sites are written inside each *substrate circle*. Below the *substrate circles* are given the expected fragment lengths following *Pst*I or *Eco*RI digestion. The two circular products of intramolecular recombination between *attP* and *attB* are drawn at the *bottom left*. They are shown separated from one another but after a typical recombination reaction they are linked to one another as a catenane (21). The single dimeric circle that arises from recombination between *attP* on one circle with *attB* on the other is drawn at the *lower right*. Restriction sites, attachment sites, and fragment lengths are indicated as for the substrate.

In addition, we have investigated this reaction in more detail. Optimal conditions for IHF-independent recombination promoted by Int-h are slightly different than those observed for recombination mixtures in which Int-h or Int$^+$ are

supplemented by IHF. The optimal ionic strength is lower (25 mM KCl rather than 70 mM), the kinetics are about 2-fold slower (half-maximal recombination in 30 min rather than 10–20 min), and more Int protein is required for maximal recombination (about 2-fold as much). However, even under these optimal conditions, in the absence of IHF, Int-h does not recombine more than 15% of the substrate DNA.

The DNA substrate used in Table II and Fig. 3 contains a pair of attachment sites oriented as a direct repeat. Integrative recombination promoted by Int-h in the absence of IHF has been observed with two other kinds of substrates. One substrate, pBP90, contains *attP* and *attB* on the same circle of DNA with the two attachment sites oriented so their core sequences form an inverted repeat. The other substrate was a pair of circles each of which carries a single attachment site, either *attP* or *attB*. Both substrates exhibited approximately the same capacity for recombination promoted by Int-h in the absence of IHF (measured relative to recombination in the presence of IHF) as was observed for pBP86 substrate (Table III and data not shown). Taken together, these results show that the capacity of Int-h to carry out recombination in the absence of IHF depends neither on the number of sites per circle nor on their orientation (see Ref. 20 for an unusual case of λ site-specific recombination that strongly depends on these factors).

Could recombination with Int-h in the absence of IHF reflect the action of a second protein that contaminates our purified Int preparation? This putative contaminant cannot be normal IHF because the Int-h protein was purified from a strain partially deleted for the *himA* gene. In addition, SDS gel electrophoresis reveals no trace of a polypeptide with the mobility of the remaining subunit of IHF (Fig. 2), even when the gel is greatly overloaded (data not shown). Since any putative contaminant would have to be present in very small amounts relative to Int, it would have to function catalytically, a mode of action not observed with IHF. We attempted to test for such an activity in the following way. We inactivated Int-h protein by treatment with *N*-ethylmaleimide and assayed for residual IHF or some other *N*-ethylmaleimide-resistant component that could either activate purified Int-protein or stimulate purified Int-h protein; no such component was found. This negative result does not rule out the existence of a helping factor in our preparations of Int-h, but the simplest interpretation of our results is that Int-h protein has an intrinsic capacity to carry out all the steps required for integrative recombination.

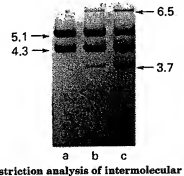


FIG. 5. **Restriction analysis of intermolecular *versus* intramolecular recombination in the presence or absence of IHF.** Recombination of supercoiled pBP86 substrate was carried out with the following concentrations of IHF and Int: *lane a*, 0.0 μg/ml IHF and 0.0 μg/ml Int; *lane b*, 0.0 μg/ml IHF and 6.0 μg/ml Int-h; *lane c*, 0.83 μg/ml IHF and 6.0 μg/ml Int-h. Reactions were carried out in the presence of 25 mM KCl for 60 min at 25 °C, treated with *Pst*I restriction endonuclease, and electrophoresed as described under "Experimental Procedures." The position of the substrate fragments is indicated at the *left* and the positions of the fragments diagnostic for intermolecular recombination (6.5 kb) and total recombination (3.7 kb) are indicated at the *right*.

Although many features of IHF-independent recombination sponsored by Int-h are similar to those of standard integrative recombination, there is one surprising difference. IHF-independent recombination preferentially recombines attachment sites on different molecules as opposed to sites that are situated on the same molecule. The reverse is true for integrative recombination that is promoted by either Int$^+$ or Int-h when supplemented by IHF. The bias favoring intermolecular *versus* intramolecular recombination is revealed by restriction analysis. Fig. 4 shows restriction maps for pBP86, a substrate with directly repeated attachment sites; recombination between *attP* and *attB* leads to the appearance of three kinds of new bands after digestion with restriction endonuclease *Pst*I: those that arise strictly from intramolecular recombination (1.4-kb circle), those that come strictly from intermolecular recombination (6.5-kb linear fragment), and those that are produced by both pathways (3.7-kb linear fragment).[3] Fig. 5 shows that when recombination reactions carried out in the presence or absence of IHF are analyzed by restriction with *Pst*I nuclease, the two linear recombinant fragments are produced in different relative yields. In reactions carried out in the presence of IHF (*lane c*), the 6.5-kb fragment is much less prominent than the 3.7-kb fragment, indicating that intermolecular recombination plays only a small role and that intramolecular is the favored reaction. Intramolecular recombination is also the dominant pathway in reaction mixtures with wild-type Int and IHF (data not shown), supporting earlier conclusions (11, 21). In contrast, in reactions carried out in the absence of IHF (*lane b*), the 6.5-kb fragment and 3.7-kb fragment are of similar intensity, indicating that intermolecular recombination is a major pathway. When the amount of DNA in each band is quantitated (Table III), it appears that essentially all of the recombinant product comes from the intermolecular pathway in the absence of IHF and that less than 5% of the recombinant product comes from this route in the presence of IHF. This analysis is supported by the appearance of a substantial amount of 1.4-kb circular species in reactions carried out in the presence of IHF and the virtual absence of such species in reactions carried out in the absence of IHF (data not shown). The same kind of analysis of intermolecular *versus* intramolecular recombination has been done with two other substrates: pBP1, that is similar to pBP86 but has a different separation between attachament sites (6.8 *versus* 1.4 kb) and pBP90, that has *attP* and *attB* oriented as an inverted repeat. Equivalent results were obtained; in the absence of IHF, recombination promoted by Int-h was predominantly intermolecular.

The conclusion that, in the absence of IHF, Int-h mainly promotes intermolecular recombination is confirmed by an analysis of the unrestricted products of reaction mixtures. To avoid complications of supercoiling, the products of recombination of pBP86 substrate were nicked with pancreatic DNase and electrophoresed on a high resolution gel system as first described by Sundin and Varshavsky (22). Fig. 6, *lanes b* and *c*, show that recombination in the presence of IHF yields a

[3] Note that restriction with endonuclease *Eco*RI yields only fragments that are of the third kind. Thus, a bias in intermolecular *versus* intramolecular recombination would not have been detected in the standard assay used to purify Int-h. Note also that recombination between *attP* one molecule with *attP* on another does not change the restriction pattern and is not scored in these assays.

TABLE III

Quantitative comparison of recombination pathways

Recombination was carried out as described in the legend to Fig. 5 and analyzed by treatment with endonuclease *Pst*I and agarose gel electrophoresis, as described under "Experimental Procedures." The substrate pBP86 and its recombinant products are described in Fig. 4. Recombination between DNA molecules that contain only a single attachment site was carried out with supercoiled plasmid pPA1 (containing *attP*) and *Eco*RI-linearized plasmid pBB105 (containing *attB*). Treatment of these reaction mixtures with endonuclease *Pst*I yields substrate fragments of 4.2 and 4.9 kb (plus smaller fragments) and a recombinant fragment of 7.2 kb (plus smaller fragments). wt, wild type protein; h, Int-h protein.

Substrate	IHF	Int	Recombination	
			Total	Intermolecular
			%	% *total recombination*
pBP86	+	wt	66.0[a]	2.0[b]
pBP86′	+	h	63.2[a]	4.7[b]
pBP86′	−	h	4.6[a]	113.0[b]
pPA1 × pBB105	+	wt	29.8[d]	(100)[e]
pPA1 × pPB105′	+	h	25.3[d]	(100)[e]
pPA1 × pBB105′	−	h	1.6[d]	(100)[e]

[a] The cpm in the 3.7-kb fragment was divided by the cpm in the 4.4 plus 5.1-kb fragments from an unrecombined mixture. This ratio was multiplied by the factor 2.57 to correct for the difference in fragment sizes; the result is expressed as a percentage.

[b] The cpm in the 6.5-kb fragment was divided by the cpm in the 4.4 plus 5.1 kb fragments from an unrecombined mixture. This ratio was multiplied by the factor 1.45 to correct for the difference in fragment sizes. This value, the fraction of substrate recombining via the intermolecular pathway, was divided by the total recombinant fraction to give the proportion of recombinants that use the intermolecular pathway. This result is expressed as a percentage.

[c] Average of four experiments.

[d] The cpm in the 7.2-kb fragment was divided by the cpm in the 4.2 plus 4.9 kb fragments from an unrecombined mixture. This ratio was multiplied by 1.265 to correct for the difference in fragment sizes. The result is expressed as a percentage.

[e] For these substrates, intermolecular recombination is the only pathway possible.

[f] Average of three experiments.

ladder of bands. Each band in the ladder contains catenanes between the two circular products of intramolecular recombination (21), each step of the ladder representing catenanes with a different number of interlocks between the two circles (22). In contrast, Fig. 6*d* shows that in the absence of IHF, recombination promoted by Int-h yields mostly dimeric circles. Isolation of the dimer band and subsequent restriction with endonuclease *Eco*RI confirmed that these circles are the result of intermolecular recombination between an *attP* on one substrate molecule and an *attB* on a second molecule (data not shown).

Recombination of Nonsupercoiled DNA—The *int-h* mutation has been reported to increase λ site-specific recombination during infection of strains carrying a mutation in *gyr*B, the gene encoding the β subunit of DNA gyrase (9). This suggests that Int-h protein might be more active than Int⁺ on DNA substrates that have reduced levels of supercoiling. Fig. 7 shows a comparison of recombination with supercoiled and nonsupercoiled substrates. The assays were carried out in the presence of IHF at low ionic strength, a condition that is optimal for Int⁺-promoted recombination of nonsupercoiled substrates. As previously reported (23, 24), even under these conditions supercoiled DNA is the better substrate for wild-type Int protein, recombining about 10 times as fast as nonsupercoiled DNA (Fig. 7*a*). Int-h protein is similar to Int⁺ in the speed with which it promotes recombination of supercoiled DNA substrates (Fig. 7*b*). However Int-h promotes recombination of nonsupercoiled DNA with an initial velocity about 3 times faster than that observed with Int⁺. We conclude from these experiments that, when supplemented by IHF, Int-h is

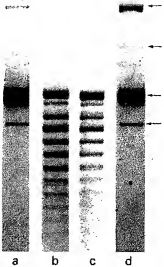


FIG. 6. **Topological analysis of intermolecular *versus* intramolecular recombination.** Supercoiled plasmid pBP86 was recombined as described in the legend to Fig. 5 with the following concentrations of IHF and Int: *lane a*, 0.0 μg/ml IHF and 0.0 μg/ml Int; *lane b*, 0.83 μg/ml IHF and 4.0 μg/ml Int⁺; *lane c*, 0.50 μg/ml IHF and 1.5 μg/ml Int-h; *lane d*, 0.0 μg/ml IHF and 6.0 μg/ml Int-h. After 60 min at 25 °C, the reaction was adjusted to 100 mM KCl, 10 mM $MgCl_2$ and digested for 10 min at 37 °C with sufficient pancreatic DNase to introduce several nicks per substrate. The samples were then electrophoresed as described (22). The position of the following monomeric (9.4 kb) and dimeric species are indicated: *1*, linear monomer; *2*, nicked monomer circle; *3*, linear dimer; *4*, nicked dimer circle. The linear species arise from excessive digestion with pancreatic DNase.

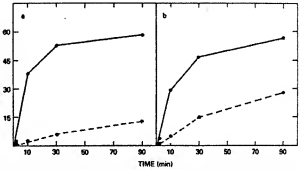


FIG. 7. **Recombination of supercoiled and nonsupercoiled substrates.** The substrate was a mixture of two DNAs, the supercoiled form of pBP86 and the nicked circle form of pBP86Δ1, a deletion derivative of pBP86 (constructed in this laboratory by T. Pollock) that has lost 600 base pairs from the 1.4-kb segment that separates *attP* and *attB*. An aliquot of this substrate (0.5 μg) was incubated as described in the legend to Fig. 5 with 0.5 μg/ml IHF and either 4.0 μg/ml Int⁺ (*panel a*) or 3.0 μg/ml Int-h (*panel b*). After various times, aliquots were removed and stopped by heating to 65 °C, digested with restriction endonuclease *Bam*HI, and electrophoresed through polyacrylamide. The radioactivity in a fragment diagnostic of recombination of the supercoiled substrate (1.4 kb) or nonsupercoiled substrate (0.8 kb) was determined as described under "Experimental Procedures." Each value was divided by the radioactivity in a *Pst*I/*Bam*HI fragment characteristic of the appropriate substrate and the ratio was adjusted to yield the extent of recombination as in Table III. The average of four and six experiments is plotted in *a* and *b*, respectively. ——, supercoiled substrate; – – –, nonsupercoiled substrate.

superior to Int$^+$ in promoting recombination of nonsupercoiled DNA.

The results just presented show that supercoiling imparts a modest benefit to recombination promoted by Int-h. We next asked whether this benefit requires the presence of IHF. Supercoiled and nonsupercoiled DNA were compared as substrates for Int-h promoted recombination in the absence of IHF. Inspecting gels like those used to generate Fig. 7 indicated that the efficiency and kinetics of recombination were similar for both substrates (data not shown). This impression was confirmed by an experiment in which the extent of recombination was quantitated. After 20 min of incubation, supercoiled and nonsupercoiled substrates both had undergone 3.5% recombination; after 90 min, the yield of recombinants from both substrates was 14.0%. It appears that Int-h protein, by itself, does not distinguish between supercoiled and nonsupercoiled DNA substrates. We have considered one trivial explanation for this lack of discrimination. It has been shown earlier that Int protein contains a topoisomerase activity that relaxes supercoiled DNA (4). This activity is intrinsically weak relative to the recombination activity of Int and is partially inhibited by IHF (25). If Int-h protein displayed a much stronger topoisomerase activity, supercoiled DNA might be relaxed in reaction mixtures before recombination could occur. However, this hypothesis is not supported by a comparison of the relaxing activities of Int$^+$ and Int-h proteins (Fig. 8). Moreover, no major difference is seen in the relaxation of the substrate DNA when similar amounts of Int$^+$ and Int-h are used in recombination reactions (data not shown). We conclude that the failure of Int-h to recombine supercoiled and nonsupercoiled substrates with different efficiency means that IHF is an essential part of the mechanism that senses the superhelicity of the recombination substrate.

DISCUSSION

Purified Int-h protein carries out substantial amounts of integrative recombination in the apparent absence of IHF. Thus, one must conclude that Int protein can manifest all the activities required for the elementary steps of recombination between specific sites. As mentioned in the Introduction, earlier observations indicated that wild-type Int carries the catalytic center responsible for breakage and reunion. Our studies with Int-h confirm these conclusions and demonstrate, for the first time, that this protein can also specify the steps required for synapsis. Unless the *int-h* allele creates new properties rather than enhancing properties inherent in the parent gene, we must conclude that wild-type Int also has the capacity to carry out synapsis as well as recognition and strand exchange. Indeed, when sensitive methods are used to probe for recombinant products, wild-type Int does reveal some ability to carry out recombination in the absence of IHF, albeit at a level 50-fold lower than that observed with Int-h acting alone and 500-fold lower than observed in the presence of IHF (Table II). By implication, IHF must play an accessory role in recombination, enhancing the capacity of Int to carry out one or more of the critical steps in recombination.

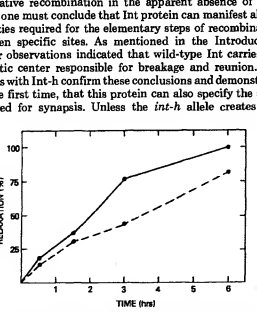


FIG. 8. **Relaxing activity of Int$^+$ and Int-h.** Radioactively labeled supercoiled plasmid pPA1 (DNA) was incubated as described under "Experimental Procedures" with identical amounts of Int$^+$ (——) or Int-h (– – –). The reactions were analyzed by cesium chloride-ethidium bromide centrifugation as described under "Experimental Procedures." The degree of relaxation relative to a fully relaxed plasmid is plotted as a function of the length of the incubation.

What mechanism underlies the enhanced capacity of Int-h to carry out recombination in strains mutant for IHF? Our studies have ruled out three kinds of explanation. First, we have shown that the *int-h* allele does not simply lead to the overproduction of an essentially wild-type protein. The amount of Int protein, we recover from *him*A cells expressing Int$^+$ or Int-h alleles is virtually identical, confirming and extending studies on the rate of synthesis of Int$^+$ and Int-h polypeptides (9). A second possible explanation for the enhanced activity of the *int-h* allele is that it specifies an Int protein with an enhanced capacity for enzymatic turnover. Previous studies had demonstrated that many molecules of purified Int$^+$ protein are required to produce a single recombinant (15, 26, 27). This implies that Int promotes recombination by a stoichiometric rather than catalytic action on attachment sites. If a molecule of Int-h protein were able to participate in more than one recombination event, an intrinsically weak capacity to carry out recombination in the absence of IHF would be magnified. However, the same value for the amount of protein required to produce a recombinant is observed with our purified preparation of Int-h protein as is seen with purified Int$^+$. Therefore, an increased capacity for turnover is not the basis for the Int-h phenotype. A third hypothesis for the behavior of Int-h invokes preferential interaction between Int-h and IHF, permitting Int-h to utilize IHF whose quality or quantity is altered by mutation. This hypothesis, which was our favorite at the outset of this work, is ruled out as the sole explanation for the behavior of Int-h by our observation that Int-h carries out recombination in the complete absence of IHF. Titration curves show about a 3-fold difference in the levels of IHF required to stimulate recombination by Int-h and Int$^+$. This small effect and the enhanced recombination in *him*A42 (as opposed to other IHF mutants) indicate that Int-h may have a preferred interaction with IHF but this putative alteration cannot be a major factor in explaining the Int-h phenotype.

Many possibilities remain as plausible explanations for the behavior of Int-h protein. These can be organized around the concept that recombination can be divided into steps of recognition, synapsis, and strand exchange. Alteration in recognition of attachment sites by Int-h is a very straightforward hypothesis. For example, Int-h might have a higher affinity for the core region of *attP* than does Int$^+$. Footprinting studies have shown that IHF binds to segments of *attP* that flank the region of the core that is occupied by Int (reviewed in Ref. 1). Moreover, similar studies have revealed that IHF enhances the binding of Int$^+$ to the core region.[4] Thus, it is attractive to imagine that Int-h can function in the absence of IHF because it binds more tightly to the core than does Int$^+$. This argument can easily be extended to explain the enhanced capacity of Int-h to promote integration at secondary bacterial attachment sites (9). In addition to models invoking altered

[4] N. Craig and H. Nash, manuscript in preparation.

recognition of attachment sites, equally attractive hypotheses can be constructed concerning alterations in the capacity of Int-h to carry out the later steps of recombination. For this purpose, it is best to consider a formal scheme for synapsis and strand exchange embodied by the equation:

$$attP^* + attB^* \rightleftharpoons (attP^*\text{-}attB^*) \rightarrow attL + attR$$

In this scheme, attP* and attB* represent attachment sites loaded with recombination proteins, (attP*-attB*) represents the synaptic intermediate in which the two sites are juxtaposed with the 15-base pair cores aligned, and attL + attR represent the products of strand exchange. Since synapsis is postulated to be a reversible process, the efficiency of recombination will be enhanced by changes that either stabilize the synaptic intermediate or accelerate its conversion to a recombinant product. If the alteration in Int-h changes either of these characteristics, it would be easy to understand how this protein might be better able to utilize what little synaptic intermediate might form under restrictive conditions such as the absence of IHF. Although our experiments have not yet defined which of these mechanisms is responsible for the altered recombination activity of Int-h, we feel the present work opens the way to a rational investigation of this problem.

Our most surprising finding is that, when acting alone, Int-h preferentially recombines two attachment sites that are located on separate circles rather than two sites that are situated on the same circle. That is to say, in the absence of IHF, Int-h promotes intermolecular rather than intramolecular recombination. Precisely the opposite is true when Int-h carries out recombination in the presence of IHF. This same bias, intramolecular recombination over intermolecular recombination, has been observed both *in vivo* (12) and *in vitro* (21) for wild-type Int protein in the presence of IHF. A preference for intramolecular recombination is readily understandable since the effective concentration of attachment site pairs should be high when the sites are tethered to each other by a length of flexible DNA (28, 29). Thus, the preference for intermolecular recombination shown by Int-h in the absence of IHF is surprising. This bias has been observed with two different substrates: one containing directly repeated attachment site and one containing an inversely repeated pair of sites. We think, therefore, that the phenomenon is intrinsic to IHF-independent recombination and we have considered two kinds of explanation for our observations. The first model states that, in the absence of IHF, intramolecular recombination is suppressed while intermolecular recombination proceeds at its normal rate. This could come about as a result of Int-h binding to substrate DNA. It is known that, in addition to specific binding to attachment sites, Int protein has a substantial nonspecific affinity for DNA. Footprinting (6) and electron microscopic (10) studies have shown that, at high ratios of Int to DNA, long stretches of DNA become covered with protein. If this were to happen to a recombination substrate, the DNA between attachment sites might be made sufficiently stiff so that the capacity of two sites on the same circle to becomed juxtaposed would decrease. A second class of models to explain the bias toward intermolecular recombination has been suggested by the observation that Int can aggregate DNA.[5] If the local concentration of DNA molecules is raised sufficiently by aggregation, statistical theory implies that random intermolecular collisions between attachment sites will predominate over intramolecular events (28, 29). An interesting example of this kind of phenomenon has been recently observed for the joining of cohesive ends by DNA ligase in the presence or absence of volume excluders like polyethylene glycol. In the absence of polyethylene glycol, intramolecular ligation to form circles is the predominant mode but when the effective concentration of DNA is raised by the addition of polyethylene glycol, circles are not formed and linear multimers accumulate (30). We imagine that, in the absence of IHF, Int-h may promote aggregation of substrate DNA either because of nonspecific charge-shielding effects or because of interactions between Int proteins bound to the DNA. Regardless of which proposed mechanism is responsible for the preference for intermolecular recombination, it should be emphasized that IHF reverses this bias. This means that IHF not only stimulates recombination promoted by Int-h and Int$^+$ but also changes the capacity of Int to bind nonspecifically to DNA and/or to form DNA aggregates. In this context, it is interesting to note that IHF prevents the formation of DNA aggregates by Int.[5]

IHF may not be the only *E. coli* protein that can reverse the intermolecular recombination bias of Int-h. Substantial amounts of intramolecular recombination are observed when Int-h is supplemented with crude extracts from cells carrying a deletion in the *himA* gene.[6] It may be that many basic proteins will share with IHF the capacity to interfere with non-specific binding to DNA or aggregation of DNA by Int-h. We are led to speculate that the action of many DNA binding proteins that, like Int, have a significant nonspecific binding affinity and a tendency to aggregate is modulated by the kind of interaction with accessory proteins that we have uncovered in this work.

Acknowledgments—We thank Nancy Craig for discussion during the course of this work, Paul Kitta, Evelyne Richet, and Robert Weisberg for their comments on this manuscript, and Arthur Landy for communicating results prior to publication. We are grateful to Carol Robertson for expert technical assistance and to Marge Schnackenberg for preparing this manuscript.

REFERENCES

1. Weisberg, R. A., and Landy, A. (1983) in *Lambda II* (Hendrix, R. W., Roberts, J. W., Stahl, F. W., and Weisberg, R. A., eds) pp. 211–250, Cold Spring Harbor Laboratory, NY
2. Ross, W., Landy, A., Kikuchi, Y., and Nash, H. (1979) *Cell* **18**, 297–307
3. Kikuchi, Y., and Nash, H. A. (1979) *Proc. Natl. Acad. Sci. U. S. A.* **76**, 3760–3764
4. Craig, N. L., and Nash, H. A. (1983) *Cell* **35**, 795–803
5. Hsu, P.-L., and Landy, A. (1984) *Nature (Lond.)*, in press
6. Ross, W., and Landy, A. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79**, 7724–7728
7. Hsu, P.-L., Ross, W., and Landy, A. (1980) *Nature (Lond.)* **285**, 85–91
8. Ross, W., and Landy, A. (1983) *Cell* **33**, 261–272
9. Miller, H. I., Mozola, M. A., and Friedman, D. I. (1980) *Cell* **20**, 721–729
10. Better, M., Lu, C., Williams, R. C., and Echols, H., (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79**, 5837–5841
11. Pollock, T. J., and Nash, H. A. (1983) *J. Mol. Biol.* **170**, 1–18
12. Nash, H. A. (1974) *Virology* **57**, 207–216
13. So, M., Gill, R., and Falkow, S. (1975) *Mol. Gen. Genet.* **142**, 239–249
14. Honigman, A., Hu, S-L., and Szybalski, W. (1979) *Virology* **92**, 542–556
15. Nash, H. A., and Robertson, C. A. (1981) *J. Biol. Chem.* **256**, 9246–9253
16. Mizuuchi, K., Fisher, L. M., O'Dea, M. H., and Gellert, M. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **4**, 1847–1851
17. Nash, H. A. (1983) *Methods Enzymol.* **100**, 210–216
18. Kikuchi, Y., and Nash, H. (1979) *Cold Spring Harbor Symp. Quant. Biol.* **43**, 1099–1109

[5] D. Straney and A. Landy, personal communication.

[6] B. Lange-Gustafson and H. Nash, unpublished observation.

19. Shimada, K., and Campbell, A. (1974) *Proc. Natl. Acad. Sci. U. S. A.* **71**, 237–241
20. Craig, N. L., and Nash, H. A. (1983) in *Mechanisms of DNA Replication and Recombination* (Cozzarelli, N., ed) pp. 617–636, Alan R. Liss, Inc., New York
21. Mizuuchi, K., Gellert, M., Weisberg, R. A., and Nash, H. A. (1980) *J. Mol. Biol.* **141**, 485–494
22. Sundin, O., and Varshavsky, A. (1980) *Cell* **21**, 103–114
23. Pollock, T. J., and Abremski, K. (1979) *J. Mol. Biol.* **131**, 651–654
24. Pollock, T. J., and Nash, H. A. (1983) in *Genetic Rearrangement* (Chater, K. F., Cullis, C. A., Hopwood, D. A., Johnson, A. W. B., and Woolhouse, H. W., eds) pp. 41–57, Croom Helm, London
25. Nash, H. A., Mizuuchi, K., Enquist, L. W., and Weisberg, R. A. (1981) *Cold Spring Harbor Symp. Quant. Biol.* **45**, 427–428
26. Nash, H. A. (1981) in *The Enzymes* (Boyer, P. D., ed) Vol. XIV, pp. 471–480, Academic Press, New York
27. Kikuchi, Y., and Nash, H. A. (1978) *J. Biol. Chem.* **253**, 7149–7157
28. Wang, J. C., and Davidson, N. (1966) *J. Mol. Biol.* **19**, 469–482
29. Shore, D., Langowski, J., and Baldwin, R. L. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78**, 4833–4837
30. Zimmerman, S. B., and Pheiffer, B. H. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80**, 5852–5856
31. Hoess, R. H., Foeller, C., Bidwell, K., and Landy, A. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77**, 2482–2486
32. Court, D., Schmeissner, U., Bear, S., Rosenberg, M., Oppenheim, A. B., Montanez, C., and Guarneros, G. (1983) in *Gene Expression* (Hamer, D. H., and Rosenberg, M. J., eds) pp. 311–326, Alan R. Liss, Inc., New York

EXHIBIT 6

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:
Peter DROGE, Nicole CHRIST and Elke LORBACH

Serial No.: 10/082,772

Filed: February 22, 2002

For: SEQUENCE-SPECIFIC DNA RECOMBINATION IN EUKARYOTIC CELLS

Group Art Unit: 1636

Examiner: Q. Nguyen

Atty. Dkt. No.: DEBE:008US/SLH

CERTIFICATE OF ELECTRONIC SUBMISSION

DATE OF SUBMISSION: October 4, 2006

DECLARATION OF PETER DRÖGE UNDER 37 C.F.R. §1.132

I, Dr. Peter Dröge, do declare that:

1. I am a citizen of Germany residing in Singapore. I currently hold the position of Associate Professor and Head of Division at the School of Biological Sciences, Nanyang Technological University, Singapore. My research experience includes genetic modification technologies and molecular genetics. A copy of my *curriculum vitae* is attached.

2. I have reviewed the above-captioned application, a copy of the pending claims, and the Office Action mailed on May 4, 2006, as well as the literature cited therein. I understand that the examiner in charge of the present application asserts that one of ordinary skill in the art would find the present invention "obvious" over the citations of Hartley and Christ & Dröge, and Crouzet and Christ & Dröge, optionally in view of Capecchi. For the following reasons, I respectfully disagree.

3. The Int-h/218 mutant was originally generated with an aim to design a recombinase which exhibits an increased binding affinity for so-called core binding sites. The latter are present in all *att* sites (the recombination substrates). This enzyme (and the parental mutant Int-h) has never been studied in detail *in vitro*, *i.e.*, purified and analyzed with DNA substrates in the test tube. It was, therefore, not clear whether the enzyme would be active in the absence of protein co-factors and negative DNA supercoiling of substrate DNA. However, both factors are present in *E. coli*. Before we transferred this mutant to mammalian cells, we knew that the enzyme could catalyze an abnormal reaction in *E. coli* in the absence of the co-factor IHF (Christ and Dröge, 1999). However, one has to realize that DNA substrates (whether episomal or genomic) are negatively supercoiled inside *E. coli*. It was, therefore, not obvious to one of ordinary skill to deduce from the existing data that the mutant recombinase would work inside mammalian cells where the DNA is topologically relaxed. In fact, up to this day, the reason why both Int-h and the double mutant Int-h/218 are functional in eukaryotic cells remains a mystery. One possibility is that there is an unidentified mammalian co-factor which supports the prokaryotic recombinase. Based on these facts, a claim that the invention is "obvious" reflects a thorough misunderstanding of the topic.

4. I declare that all statements made herein of my own knowledge are true, and that all statements of my own belief are believed to be true, and further that these statements were made with the knowledge that willful false statements are punishable by fine or imprisonment, or both, under § 1001 of title 18 of the United States Code, and that such willful false statements may jeopardize the validity of this patent, and any reexamination certificate issuing thereon.

Sep - 29th, 2006
Date

Dr. Peter Dröge

EXHIBIT 7

US006632672B2

(12) **United States Patent**
Calos

(10) **Patent No.: US 6,632,672 B2**
(45) **Date of Patent: *Oct. 14, 2003**

(54) **METHODS AND COMPOSITIONS FOR GENOMIC MODIFICATION**

(75) Inventor: **Michele P. Calos**, Woodside, CA (US)

(73) Assignee: **The Board of Trustees of the Leland Stanford Junior University**, Palo Alto, CA (US)

(*) Notice: This patent issued on a continued prosecution application filed under 37 CFR 1.53(d), and is subject to the twenty year patent term provisions of 35 U.S.C. 154(a)(2).

Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/377,885**

(22) Filed: **Aug. 19, 1999**

(65) **Prior Publication Data**

US 2003/0050258 A1 Mar. 13, 2003

Related U.S. Application Data

(60) Provisional application No. 60/097,166, filed on Aug. 19, 1998.

(51) **Int. Cl.**[7] **C12N 15/87**; C12N 15/63; C12N 5/00

(52) **U.S. Cl.** **435/462**; 435/320.1; 435/325; 435/465

(58) **Field of Search** 536/23.1; 435/183, 435/320.1, 325, 455, 462

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,190,871	A	3/1993	Cox et al.	
5,470,727	A	11/1995	Mascarenhas et al.	
5,654,182	A	8/1997	Wahl et al.	
5,801,030	A	9/1998	McVey et al.	
5,830,698	A	11/1998	Reff et al.	435/69.1
5,888,732	A	3/1999	Hartley et al.	435/6
6,143,530	A	* 11/2000	Crouzet et al.	435/91.42

OTHER PUBLICATIONS

Kuhstoss et al. Analysis of the Integration Function of the *Streptomycete bacteriophage* φC31. J. Mol. Biol, 222, 897–907, 1991.*

AC Groth et al., "A phage integrase directs efficient site–specific integration in human cells," Dept. of Genetics, Stanford Univ. School of Med., Proc. Natl. Acad. Sci. USA May 2000,vol. 97, No. 11, pp. 5995–6000.*

MP Deonarain, "Ligand–targeted receptor–mediated vectors for gene delivery," Exp. Opin. Ther. Patents, 1998, 8:1, pp. 53–69.*

WF Anderson, "Human gene therapy," Nature, Apr. 1998, vol. 392, pp. 25–30.*

IM Verma et al., "Gene therapy–promises, problems and prospects," Nature, Sep. 1997, vol. 389, pp. 239–242.*

RG Crystal, "Transfer of Genes to Humans:Early Lessons and Obstacles to Success," Science, Oct. 1995, vol. 270, pp. 404–409.*

TM Clay et al., "Potential use of T Cell Receptor Genes to Modify Hematopoietic Stem Cells for the Gene Therapy of Cancer," Pathology Oncology Research, 1999, vol. 5, No. 1, pp. 3–15.*

FH Gage, Nature, "Cell therapy, " Apr. 1998, vol. 392,pp. 19–24.*

Thyagarajan et al. Site–specific genomic integration in mammalian cells mediated by phage C31 integrase Mol. Cell. Biol. 21, pp. 3926–3934 2001.*

Stoll et al. Phage TP901–1 site–specific integrase functions in human cells J. Bacteriol. 184(13): pp. 3657–3663 2002.*

Olivares et al. Phage R4 integrase mediates site–specific integration in human cells Gene, 278(1–2): pp. 167–176, 2001.*

Agah et al., "Targeted Expression of Cre Recombinase Provokes Cardiac–restricted, Site–specific Rearrangement in Adult Ventricular Muscle In Vivo," *J. Clin. Invest. 100*:169–179 (1997).

Albert et al., "Site–specific Integration of DNA into Wild–type and Mutant *lox* Sites Placed in the Plant Genome," *Plant Journal 7*(4):649–659 (1995).

Argos et al., "The Integrase Family of Site–specific Recombinases: Regional Similarities and Global Diversity," *EMBO J. 5*(2):433–440 (1986.

Araki et al., "Molecular and Functional Organization of Yeast Plasmic pSR1," *J. Mol. Biol. 182*:191–203 (1985).

Baubonis and Sauer, "Genomic Targeting with Purified Cre Recombinase," 21(9):2025–2029 (1993).

Broach et al., "Recombination within the Yeast Plasmid 2μ Circle is Site–Specific," *Cell 29*:227–234 (1982).

(List continued on next page.)

Primary Examiner—Ram R. Shukla
(74) *Attorney, Agent, or Firm*—Pamela J. Sherwood; Bozicevic, Field & Francis LLP

(57) **ABSTRACT**

The present invention provides methods of site-specifically integrating a polynucleotide sequence of interest in a genome of a eucaryotic cell, as well as, enzymes, polypeptides, and a variety of vector constructs useful therefore. In the method, a targeting construct comprises, for example, (i) a first recombination site and a polynucleotide sequence of interest, and (ii) a site-specific recombinase, which are introduced into the cell. The genome of the cell comprises a second recombination site. Recombination between the first and second recombination sites is facilitated by the site-specific recombinase. The invention describes compositions, vectors, and methods of use thereof, for the generation of transgenic cells, tissues, plants, and animals. The compositions, vectors, and methods of the present invention are also useful in gene therapy techniques.

38 Claims, 12 Drawing Sheets

OTHER PUBLICATIONS

Brondsted and Hammer, "Use of the Integration Elements Encoded by the Temperate Lactococcal Bacteriophage TP901–1 to Obtain Chromosomal Single–Copy Transcriptional Fusions in *Lactococcus lactis*," *Apl. & Envirol. Microbiology* 65(2):752–758 (1999).

Christiansen et al., "Characterization of the Lactococcal Temperate Phage TP901–1 and its Site–Specific Integration," *J. of Bacteriology* 176(4):1069–1076 (1994).

Christiansen et al., "A Resolvase–like Protein is Required for the Site–specific Integration of the Temperate Lactococcal Bacteriophage TP901–1," *J. of Bacteriology* 178(17):5164–5173 (1996).

Crellin and Rood, "The Resolvase/Invertase Domain of the Site–specific Recombinase TnpX is Functional and Reoganizes a Target Sequence that Resembles the Junction of the Circular Form of the *Clostridium perfringens* Transposon Tn4451," *J. of Bacteriology* 179(16):5148–5156 (1997).

Dale and Ow, "Intra– and Intermolecular Site–specific Recombination in Plant Cells Mediated by Bacteriophage P1 Recombinase," *Gene* 91:79–85 (1990).

Flotte and Carter, "Adeno–associated Virus Vectors for Gene Therapy," *Gene Therapy* 2:357–362 (1995).

Golic and Lindquist "The FLP Recombinase of Yeast Catalyzes Site–specific Recombination in the Drosophila Genome," *Cell* 59:499–509 (1989).

Golic and Golic, "Engineering the Drosophila Genome: Chromosome Rearrangements by Design," *Genetics* 144:1693–1711 (1996).

Hallet and Sherratt, "Transposition and Site–specific Recombination: Adapting DNA Cut–and–paste Mechanisms to a Variety of Genetic Rearrangements," *FEMS Microbiol. Rev* 21:157–178 (1997).

Huang et al., "Convenient and Reversible Site–specific Targeting of Exogenous DNA into a Bacterial Chromosome by Use of the FLP RecombinaseL the FLIRT System," *J. of Bacteriology* 179(19):6076–6083 (1997).

Kay et al., "Gene Therapy," *Proc. Natl. Acad Sci. USA* 94:12744–12746 (1997).

Kuhstoss and Rao, "Analysis of the Integration Function of the Streptomycete Bacteriophage ϕC31," *J. Mol. Biol.* 222:897–908 (1991).

Lyznik et al., "Activity of Yeast FLP Recombinase in Maize and Rice Protoplasts," *Nucleic Acids Res.* 21(4):969–975 (1993).

Matsuura et al., "The *sre* Gene (ORF469) Encodes a Site–specific Recombinase Responsible for Integration of the R4 Phage Genome," *J. of Bact.* 178(11):3374–3376 (1996).

Matsuzaki et al., "Chromosome Engineering in *Saccharomyces cerevisiae* by Using a Site–specific Recombination System of a Yeast Plasmid," *J. of Bacteriology* 172(2):610–618 (1990).

Muzyczka, "Use of Adeno–associated Virus as a General Transduction Vector for Mammalian Cells," *Curr. Topics Microbiol. Immunol.* 158:97–129 (1992).

Odell et al., "Site–directed Recombination in the Genome of Transgenic Tobacco," *Mol. Gen. Genet.* 223:369–378 (1990).

Rausch and Lehmann, "Structural Analysis of the Actinophage ϕC31 Attachment," *Nucl. Acids. Res.* 19(19):5187–5189 (1991).

Sauer, "Multiplex Cre/Lox Recombination Permits Selective Site–specific DNA Targeting to Both a Natural and an Engineered Site in the Yeast Genome," *Nucl. Acids Res.* 24(23):4608–4613 (1996).

Sauer, "Functional Expression of the *cre–lox* Site–specific Recombination System in the Yeast *Saccharomyces cerevisiae*," *Mol. Cell. Biol.* 7(6):2087–2096 (1987).

Sauer, "Identification of Cryptic *lox* Sites in the Yeast Genome by Selection for Cre–mediated Chromosome Translocations that Confer Multiple Drug Resistance," *J. Mol. Biol.* 223:911–928 (1992).

Sauer and Henderson, "Targeted Insertion of Exogenous DNA into the Eukaryotic Genome by the Cre Recombinase," *New Biologist* 2:441–449 (1990).

Shirai et al., "Site–specific Integration of the Actinophage R4 Genome into the Chromosome of *Streptomyces parvulus* Upon Lysogenization," *J. of Bacteriology* 173(13):4237–4239 (1991).

Sternberg and Hamilton, "Bacteriophage P1 Site–specific Recombination, I. Recombination Bewteen *loxP* Sites," *J. Mol. Biol.* 150:467–486 (1981).

Sternberg et al., "Bacteriophage P1 Site–specific Recombination, II. Recombination Between *loxP* and the Bacterial Chromosome," *J. Mol. Biol.* 150:487–507 (1981).

Sternberg, "Bacteriophage P1 *cre* Gene and its Regulatory Region, Evidence for Multiple Promoters and for Regulation by DNA Methylation," *J. Mol. Biol.* 187197–212 (1986).

Thorpe and Smith, "In Vitro Site–specific Integration of Bacteriophage DNA Catabyzed by a Recombinase of the Resolvase/Invertase Family," *PNAS USA* 95:5505–5510 (1998).

van Deursen et al., "Cre–mediated Site–specific Translocation Between Nonhomologous Mouse Chromosomes," *Natl. Acad. Sci. USA* 92:7376–7380 (1995).

Verma and Somia, "Gene Therapy—Promises, Problems and Prospects," *Nature* 389:239–242 (1997).

Buchholz et al., "Improved Properties of FLP Recombinase Evolved by Cycling—Mutagenesis," *Nature Biotechnology* 16:657–662 (1998).

Christiansen et al., "A Resolvase–Like Protein is Required for the Site–Specific Integration of the Temperate Lactococcal Bacteriophage TP901–1," *Journal of Bacteriology* 178(17):5164–5173 (1996).

Dorgai et al., "Identifying Determinants of Recombination Specificity: Construction and Characterization of Mutant Bacteriophage Integrases," *J. Mol. Biol.* 252:178–188 (1995).

Hartung et al., "CRE Mutants with Altered DNA Binding Properties," *Journal of Biological Chemistry* 273:22884–22891 (1998).

Matsuura et al., "The sre Gene (ORF469) Encodes a Site–Specific Recombinase Responsible for Integration of the R4 Phage Genome," *Journal of Bacteriology* 178(11):3374–3376 (1996).

Thorpe et al., "In Vitro Site–Specific Integration of Bacteriophage DNA Catalyzed by a Recombinase of the Resolvase/Invertase Family," *Proc. Natl. Acad. Sci. U.S.A.* 95:5505–5510 (1998).

* cited by examiner

FIG. 1A

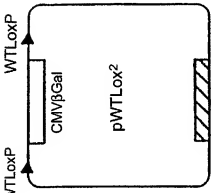


FIG. 1B

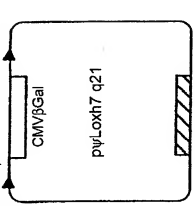


FIG. 1C

WTLoxP	ATAACTTCGTATA ATGTATGC TATACGAAGTTAT	(SEQ ID NO: 20)
ψLoxh7q21	ATACATACGTATA TATGTATA TATACATATATAT	(SEQ ID NO: 21)
ψCoreh7q21	ATAACTTCGTATA TATGTATA TATACGAAGTTAT	(SEQ ID NO: 22)

FIG. 1D

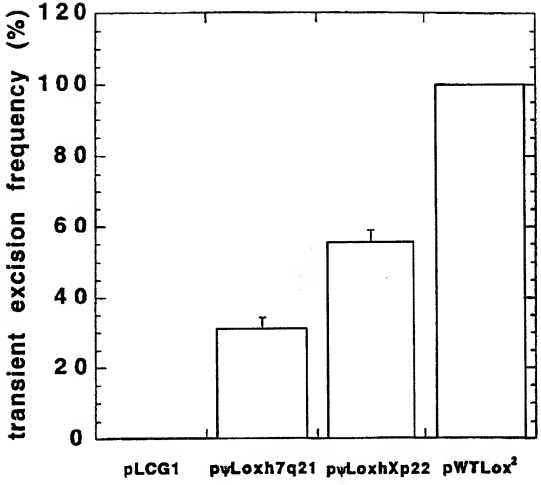

transient excision frequency (%)
120
100
80
60
40
20
0
pLCG1
pψLoxh7q21
pψLoxhXp22
pWTLox²

FIG. 2

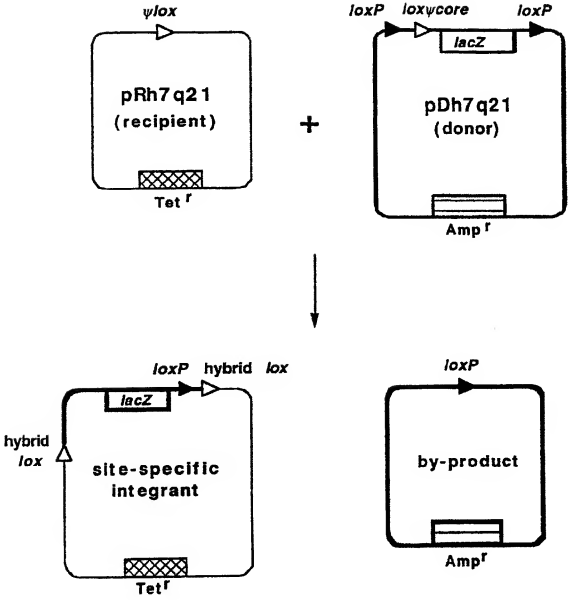

ψlox
pRh7q21
(recipient)
Tet r
+
loxP loxψcore loxP
lacZ
pDh7q21
(donor)
Amp r
loxP hybrid lox
lacZ
hybrid
lox
site-specific
integrant
Tet r
loxP
by-product
Amp r

FIG. 3

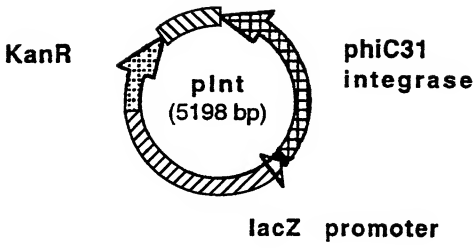

KanR
pInt
(5198 bp)
phiC31
integrase
lacZ promoter

FIG. 4A

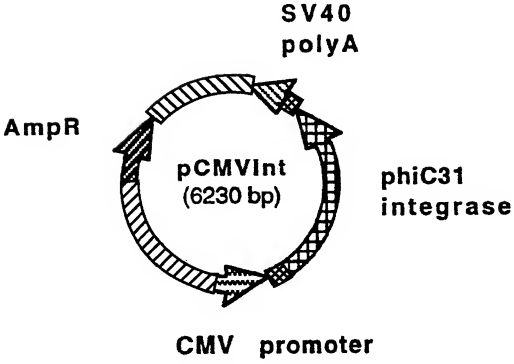

SV40
polyA
AmpR
pCMVInt
(6230 bp)
phiC31
integrase
CMV promoter

FIG. 4B

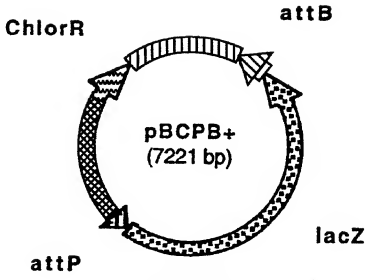

ChlorR
attB
pBCPB+
(7221 bp)
lacZ
attP

FIG. 4C

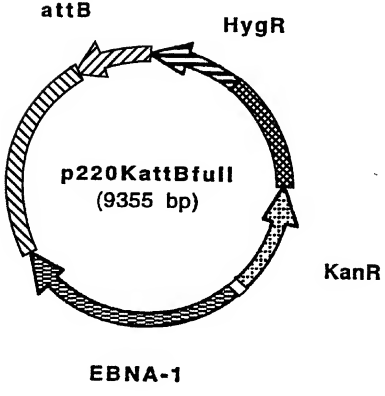

attB
HygR
p220KattBfull
(9355 bp)
KanR
EBNA-1

FIG. 4D

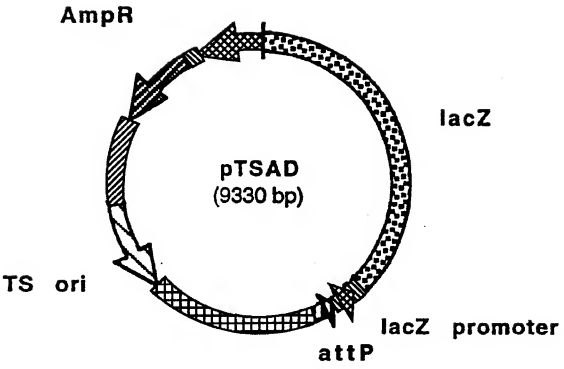


FIG. 4E

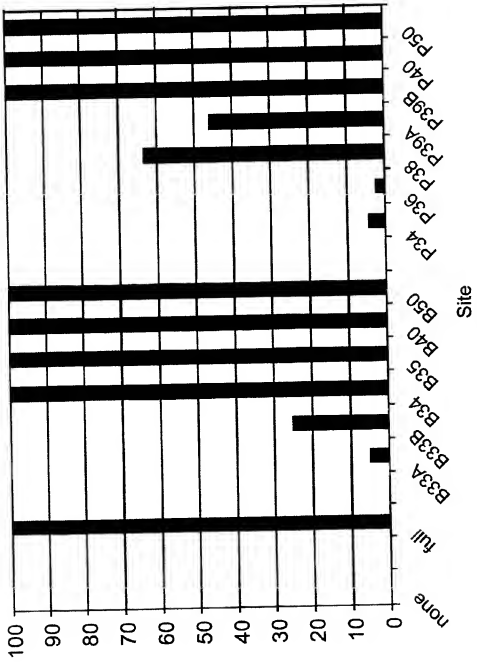
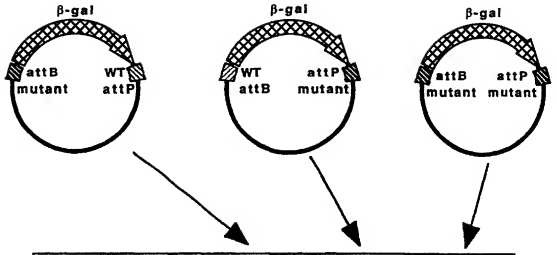

100
90
80
70
60
50
40
30
20
10
0
none
full
B33A
B33B
B34
B35
B40
B50
P34
P36
P38
P39A
P39B
P40
P50
Site

FIG. 5



MUTANT CORE WT = TTG			attB mutant	attP mutant	Both mutant
			% recombination		
A	T	G	0.9	nd	6.5
G	T	G	0	nd	1.6
C	T	G	nd	nd	100
T	***A***	G	nd	nd	nd
T	***C***	G	0.4	nd	0.2
T	***G***	G	0.1	5.1	1.5
T	T	***A***	100	100	nd
T	T	***C***	0.6	100	0
T	T	***T***	0.1	100	0.3
G	***G***	***T***	0.4	nd	0

FIG. 6

p220K			
	Blues / Total	% Integration	Corrected % Integration
no DNA	0/7,970	< .01%	< .08%
pCMVInt only	0/13,600	< .007%	< .04%
pTSAD only	1/11,830	0.008%	0.05%
pCMVInt + pTSAD	3/32,680	0.009%	0.06%
p220KattB35			
	Blues / Total	% Integration	Corrected % Integration
no DNA	0/64,530	< .002%	< .01%
pCMVInt only	0/42,720	< .002%	< .02%
pTSAD only	0/39,930	< .003%	< .02%
pCMVInt + pTSAD	382/157,710	0.242%	1.69%
p220KattBfull			
	Blues / Total	% Integration	Corrected % Integration
no DNA	0/70,350	< .001%	< .01%
pCMVInt only	0/41,960	< .002%	< .02%
pTSAD only	0/39,740	< .003%	< .02%
pCMVInt + pTSAD	1,799/ 166,890	1.08%	7.50%

FIG. 7

Wild-type *loxP*

(SEQ ID NO: 20) ATAACTTCGTATA **ATGTATGC** TATACGAAGTTAT

		(LAB DESIGNATION)	Accession number	Location/ gene
Human				
(SEQ ID NO: 23)	ACAACCATTTATA **ATATATAA** TATATGATGTTAT	PSTR1	HSAC002082	Xp22
(SEQ ID NO: 24)	ATACATACGTATA **TATGTATA** TATACATATATAT	P2	AC002383	7q21
(SEQ ID NO: 25)	ATATACACGTATA **TATATATA** TATACGTATATAT	P4	Z98048	22q13
(SEQ ID NO: 26)	CAAACAAGGTATA **TGCCTGTA** TATACGAAATGGT	P3	AC003081	7q31
(SEQ ID NO: 27)	ATATATACGTATA **TATACATA** TATACGTATATAT	Bh1	AC002090	17
(SEQ ID NO: 28)	ATATATACGTATA **TATACATA** TATACACATATAT	Bh3	HS107N3	
(SEQ ID NO: 29)	ATAAATATGTATA **TGTATATG** TATACGTATATAA	Bh4	HSAC000112	
(SEQ ID NO: 30)	ATATATATGTATA **TGTATATG** TATACGTATATAT	Bh5	HSCU90B3	
(SEQ ID NO: 31)	ATATATACGTATA **TACACATA** TATACGTATATAC	Bh7	HSXIST1	Xq13
Mouse				
(SEQ ID NO: 32)	ATATTGACATATA **TTATAAAG** TATAAGTAGTTAT	MO1	MMU28404	9 MIP1
(SEQ ID NO: 33)	GTAACTGAGTATA **TGCATATA** TATACGTATATAT	MO2	AF033025	5 MDR1
(SEQ ID NO: 34)	ATAACATATTATA **TTTATATA** TATATATATTTAA	Bm1	AC000399	9
(SEQ ID NO: 35)	ATATATATGTATA **TATATACA** TATACATACATAT	Bm2	MMBGCXD	β-globin
(SEQ ID NO: 36)	AGCACTTCCTATA **TAACTTCA** TATACGTAGCTCC	Bm3	MMU48804	PW1

FIG. 8A

		(LAB DESIGNATION)	Accession number	Location/ gene
Invertebrate				
Caenorhabditis elegans				
(SEQ ID NO: 37)	ATAGCGTCGTATA **ATCCGAAA** TATACAGATCTAT	Bi1	CEC47E12	
Plant and Fungus				
Arabidopsis thaliana				
(SEQ ID NO: 38)	CTAGTTTGGTATA **TATATATA** TATACTAATTTAT	Bp1	ATF28A23	
(SEQ ID NO: 39)	ATAACTTTGTATA **GTTTAACT** TATATTAGGTACT	Bp3	IG002P16	
Flaveria anomala				
(SEQ ID NO: 40)	ATCAGTTAGTATA **TATTCGTA** TATACGTAGATAT	Bp2	FAGDCSH	
Saccharomyces cerevisiae				
(SEQ ID NO: 41)	TTGCCTTCGTATA **TACCTTTC** TATACCAAGTAAT	By1	SCFAS1	

FIG. 8B

METHODS AND COMPOSITIONS FOR GENOMIC MODIFICATION

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is related to U.S. Provisional Patent Application Serial No. 60/097,166, filed Aug. 19, 1998, from which priority is claimed under 35 USC §119(e)(1), and which application is incorporated herein by reference in its entirety.

This invention was made with support under NIH Grant R01 DK51834 from the National Institutes of Health, U.S. Department of Health and Human Services. Accordingly, the United States Government may have certain rights in the invention.

FIELD OF THE INVENTION

The present invention relates to the field of biotechnology, and more specifically to the field of genomic modification. Disclosed herein are compositions, vectors, and methods of use thereof, for the generation of transgenic cells, tissues, plants, and animals. The compositions, vectors, and methods of the present invention are also useful in gene therapy techniques.

BACKGROUND OF THE INVENTION

Permanent genomic modification has been a long sought after goal since the discovery that many human disorders are the result of genetic mutations that could, in theory, be corrected by providing the patient with a non-mutated gene. Permanent alterations of the genomes of cells and tissues would also be valuable for research applications, commercial products, protein production, and medical applications. Furthermore, genomic modification in the form of transgenic animals and plants has become an important approach for the analysis of gene function, the development of disease models, and the design of economically important animals and crops.

A major problem with many genomic modification methods associated with gene therapy is their lack of permanence. Life-long expression of the introduced gene is required for correction of genetic diseases. Indeed, sustained gene expression is required in most applications, yet current methods often rely on vectors that provide only a limited duration of gene expression. For example, gene expression is often curtailed by shut-off of integrated retroviruses, destruction of adenovirus-infected cells by the immune system, and degradation of introduced plasmid DNA (Anderson, W F, Nature 329:25–30, 1998; Kay, et al, Proc. Natl. Acad. Sci. USA 94:12744–12746, 1997; Verma and Somia, Nature 389:239–242, 1997). Even in shorter-term applications, such as therapy designed to kill tumor cells or discourage regrowth of endothelial tissue after restenosis surgery, the short lifetime of gene expression of current methods often limits the usefulness of the technique.

One method for creating permanent genomic modification is to employ a strategy whereby the introduced DNA becomes part of (i.e., integrated into) the existing chromosomes. Of existing methods, only retroviruses provide for efficient integration. Retroviral integration is random, however, thus the added gene sequences can integrate in the middle of another gene, or into a region in which the added gene sequence is inactive. In addition, a different insertion is created in each target cell. This situation creates safety concerns and produces an undesirable loss of control over the procedure.

Adeno-associated virus (AAV) often integrates at a specific region in the human genome. However, vectors derived from AAV do not integrate site-specifically due to deletion of the toxic rep gene (Flotte and Carter, Gene Therapy 2:357–362, 1995; Muzyczk, Curr. Topics Microbiol. Immunol. 158:97–129, 1992). The small percentage of the AAV vector population that eventually integrates does so randomly. Other methods for genomic modification include transfection of DNA using calcium phosphate co-precipitation, electroporation, lipofection, microinjection, protoplast fusion, particle bombardment, or the Ti plasmid (for plants). All of these methods produce random integration at low frequency. Homologous recombination produces site-specific integration, but the frequency of such integration is very low.

Another method that has been considered for the integration of heterologous nucleic acid fragments into a chromosome is the use of a site-specific recombinase (an example using Cre is described below). Site-specific recombinases catalyze the insertion or excision of nucleic acid fragments. These enzymes recognize relatively short, unique nucleic acid sequences that serve for both recognition and recombination. Examples include Cre (Sternberg and Hamilton, J Mol Biol 150:467–486, 1981), Flp (Broach, et al, cell-29:227–234, 1982) and R (Matsuzaki, et al, J Bacteriology 172:610–618, 1990).

One of the most widely studied site-specific recombinases is the enzyme Cre from the bacteriophage P1. Cre recombines DNA at a 34 basepair sequence called loxP, which consists of two thirteen basepair palindromic sequences flanking an eight basepair core sequence. Cre can direct site-specific integration of a loxP-containing targeting vector to a chromosomally placed loxP target in both yeast and mammalian cells (Sauer and Henderson, New Biol 2:441–449, 1990). Use of this strategy for genomic modification, however, requires that a chromosome first be modified to contain a loxP site (because this sequence is not known to occur naturally in any organism but P1 bacteriophage), a procedure which suffers from low frequency and unpredictability as discussed above. Furthermore, the net integration frequency is low due to the competing excision reaction also mediated by Cre. Similar concerns arise in the conventional use of other, well-known, site-specific recombinases.

A need still exists, therefore, for a convenient means by which chromosomes can be permanently modified in a site-specific manner. The present invention addresses that need.

BRIEF DESCRIPTION OF THE INVENTION

Accordingly, in one embodiment, the present invention is directed to a method of site-specifically integrating a polynucleotide sequence of interest in a genome of a eucaryotic cell. The method comprises introducing (i) a circular targeting construct, comprising a first recombination site and the polynucleotide sequence of interest, and (ii) a site-specific recombinase into the eucaryotic cell, wherein the genome of the cell comprises a second recombination site native to the genome and recombination between the first and second recombination sites is facilitated by the site-specific recombinase. The cell is maintained under conditions that allow recombination between the first and second recombination sites and the recombination is mediated by the site-specific recombinase. The result of the recombination is site-specific integration of the polynucleotide sequence of interest in the genome of the eucaryotic cell.

The recombinase may be introduced into the cell before, concurrently with, or after introducing the circular targeting construct. Further, the circular targeting construct may comprise other useful components, such as a bacterial origin of replication and/or a selectable marker.

In certain embodiments, the recombinase may facilitate recombination between two sites designated recombinase-mediated-recombination sites (RMRS) and the RMRS comprises a first DNA sequence (RMRS5'), a core region A, and a second DNA sequence (RMRS3') in the relative order RMRS5'-core region A-RMRS3'. In this embodiment, for example, RMRS may be a loxP site or a FRT site and the recombinase may be Cre and FLP, respectively.

In additional embodiments,(i) the second recombination site is a pseudo-RMRS site, and the second recombination site comprises a first DNA sequence (attT5'), a core region B, and a second DNA sequence (attT3') in the relative order attT5'-core region B-attT3', and (ii) the first recombination site is a hybrid-recombination site comprising RMRS5'-core region B-RMRS3' or attT5'-core region B-attT3'.

In yet further embodiments, the site-specific recombinase is a recombinase encoded by a phage selected from the group consisting of ϕC31, TP901-1, and R4. The recombinase may facilitate recombination between a bacterial genomic recombination site (attB) and a phage genomic recombination site (attP), and (i) the second recombination site may comprise a pseudo-attP site, and (ii) the first recombination site may comprise the attB site or (i) the second recombination site may comprise a pseudo-attB site, and (ii) the first recombination site may comprise the attP site.

In another embodiment, (i) attB comprises a first DNA sequence (attB5'), a bacterial core region, and a second DNA sequence (attB3') in the relative order attB5'-bacterial core region-attB3', (ii) attP comprises a first DNA sequence (attP5'), a phage core region, and a second DNA sequence (attP3') in the relative order attP5'-phage core region-attP3', and (iii) wherein the recombinase meditates production of recombination-product sites that can no longer act as a substrate for the recombinase, the recombination-product sites comprising the relative order attB5'-recombination-product site-attP3' and attP5'-recombination-product site-attB3'.

In particularly preferred embodiments, (i) the second recombination site is a pseudo-attP site, the second recombination site comprises a first DNA sequence (attT5'), a core region B, and a second DNA sequence (attT3') in the relative order attT5'-core region B-attT3', (ii) the first recombination site is an attB site comprising attB5'-bacterial core region-attB3', and (iii) wherein the recombinase meditates production of recombination-product sites that can no longer act as a substrate for the recombinase, the recombination-product sites comprising the relative order attT5'-recombination-product site-attB3'{polynucleotide of interest}attB5'-recombination-product site-attT3'. Alternatively, (i) the second recombination site is a pseudo-attB site, and the second recombination site comprises a first DNA sequence (attT5'), a core region B, and a second DNA sequence (attT3') in the relative order attT5'-core region B-attT3', (ii) the first recombination site is an attP site comprising attP5'-bacterial core region-attP3', and (iii) wherein the recombinase meditates production of recombination-product sites that can no longer act as a substrate for the recombinase, the recombination-product sites comprising the relative order attT5'-recombination-product site-attP3'{polynucleotide of interest}attP5'-recombination-product site-attT3'.

In yet further embodiments, the site-specific recombinase is introduced into the cell as a polypeptide. In alternative embodiments, the site-specific recombinase in introduced into the cell as a polynucleotide encoding the recombinase and an expression cassette, optionally carried on a transient expression vector, comprises the polynucleotide encoding the recombinase.

In another embodiment, the invention is directed to a vector for site-specific integration of a polynucleotide sequence into the genome of a eucaryotic cell. The vector comprises (i) a circular backbone vector, (ii) a polynucleotide of interest operably linked to a eucaryotic promoter, and (iii) a first recombination site, wherein the genome of the cell comprises a second recombination site native to the genome and recombination between the first and second recombination sites is facilitated by a site-specific recombinase.

In certain embodiments, the recombinase normally facilitates recombination between a bacterial genomic recombination site (attB) and a phage genomic recombination site (attP) and the first recombination site may be either attB or attP.

In still another embodiment, the invention is directed to a kit for site-specific integration of a polynucleotide sequence into the genome of a eucaryotic cell. The kit comprises, (i) a vector as described above and (ii) a site-specific recombinase.

In another embodiment, the invention is directed to a eucaryotic cell having a modified genome. The modified genome comprises an integrated polynucleotide sequence of interest whose integration was mediated by a recombinase and wherein the integration was into a recombination site native to the eucaryotic cell genome and the integration created a recombination-product site comprising the polynucleotide sequence.

In certain embodiments, the recombination-site product comprises the components attT5'-recombination-product site-attB3' and attB5'-recombination-product site-attT3', wherein (i) the native recombination site is a pseudo-attP site, and the native recombination site comprises a first DNA sequence (attT5'), a core region B, and a second DNA sequence (attT3') in the relative order attT5'-core region B-attT3', (ii) the integrated polynucleotide sequence comprises a first recombination site comprising an attB site comprising attB5'-bacterial core region-attB3', and (iii) wherein the recombinase meditates production of recombination-product sites that can no longer act as a substrate for the recombinase, the recombination-product sites comprising the relative order attT5'-recombination-product site-attB3'{polynucleotide of interest}attB5'-recombination-product site-attT3'. Alternatively, the recombination-site product comprises the components attT5'-recombination-product site-attB3' and attB5'-recombination-product site-attT3', wherein (i) the native recombination site is a pseudo-attB site, and the native recombination site comprises a first DNA sequence (attT5'), a core region B, and a second DNA sequence (attT3') in the relative order attT5'-core region B-attT3', (ii) the integrated polynucleotide sequence comprises a first recombination site comprising an attP site comprising attP5'-phage core region-attP3', and (iii) wherein the recombinase meditates production of recombination-product sites that can no longer act as a substrate for the recombinase, the recombination-product sites comprising the relative order attT5'-recombination-product site-attP3'{polynucleotide of interest}attP5'-recombination-product site-attT3'.

In further embodiments, the subject invention is directed to transgenic plants and animals comprising at least one cell as described above, as well as methods of producing the same.

In yet other embodiments, the invention is directed to methods of treating a disorder in a subject in need of such treatment. The method comprises site-specifically integrating a polynucleotide sequence of interest in a genome of at least one cell of the subject, wherein the polynucleotide facilitates production of a product that treats the disorder in the subject. The site-specific integration may be carried out in vivo in the subject, or ex vivo in cells and the cells are then introduced into the subject.

A further embodiment of the invention comprises cells, tissues, transgenic animals and/or plants whose genomes have been modified using the methods described herein.

In another aspect, the present invention provides a method of modifying a genome of a cell. In the method, an attB or an attP recombination site is into the genome of a cell, wherein (i) the recombination site is recognized by a recombinase, and (ii) the cell normally does not comprise the attB or attP site. The vectors described herein and above are useful in the practice of this aspect of the invention. In a preferred embodiment, the cell that is being modified is a eucaryotic cell.

In yet another aspect, the present invention provides expression cassettes, comprising a polynucleotide encoding a site-specific recombinase, wherein (i) the recombinase is encoded by a phage (typically selected from the group consisting of φC31, TP901-1, and R4) and the recombinase is operably linked to a eucaryotic promoter. The vectors described herein and above are useful in the practice of this aspect of the invention.

These and other embodiments of the present invention will readily occur to those of ordinary skill in the art in view of the disclosure herein.

BRIEF DESCRIPTION OF THE FIGURES

FIGS. 1A through 1C are schematics of representative plasmids useful in evaluating the efficiency of pseudo-lox recombination sequences. FIG. 1A shows an unmodified plasmid containing a gene for ampicillin resistance and a gene for β-galactosidase expression (lacZ) under control of the CMV promoter (pLCG1). FIG. 1B shows the same plasmid with wild-type loxP sequences flanking the lacZ gene (pWTLox^2). FIG. 1C shows the plasmid with the ψlox h7q21 pseudo-lox recombination sequence on one side of lacZ and a lox sequence with wild-type palindromes and a pseudo-lox core on the other side (pψloxh7q21).

FIG. 1D shows the DNA sequences of the lox sites from pWTLox^2 (top line of FIG. 1D, WT LoxP (SEQ ID NO:20) and plasmid pψloxh7q21 (bottom lines of FIG. 1D, ψLoxh7q21) SEQ ID NO:21) and ψCoreh7q21 (SEQ ID NO:22)

FIG. 2 shows the results of an excision assay performed in human cells as described in the examples. Each of the tested plasmids was transfected into human 293 cells along with a Cre expression plasmid. After 72 hours, DNA was transformed into *E. coli* and recombinants scored. The transient excision frequency is expressed as a percentage, where the value for pWTLox^2 is set at 100%.

FIG. 3 is a diagram of plasmids used in a transient integration assay performed in human cells as described in the examples. pRh7q21 (upper left) was the recipient for an integration event and included the chromosomal ψlox h7q21 site (open triangle), as well as the gene for tetracycline resistance. Similar control plasmids bearing either no lox site or the wild-type loxP site were also constructed. pDh7q21 (upper right) was the donor plasmid for integration and included a lox site (open triangle, loxp|core) comprising the 8-bp core from ψlox h7q21 and the wild-type loxP palindromes. The plasmid also carried two wild-type loxP sites (dark triangles). In the presence of Cre, the plasmid origin of replication and the ampicillin resistance gene are excised, resulting in integrants that do not have two plasmid origins. This excised by-product is shown in the lower right. The site-specific integration product, bearing lacZ flanked by hybrid lox sites (shaded triangles) in a tetracycline resistant backbone, is shown at lower left. Parallel donor plasmids having, in place of ψlox h7q21, either no lox site or only wild-type loxP sites, were also constructed.

FIGS. 4A through 4E are schematic diagrams of representative plasmids used in demonstrating function of the φC31 integrase, as described in the examples. FIG. 4A shows plasmid pInt, for expression of φC31 integrase in *E. coli;* FIG. 4B depicts plasmid pCMVInt for expression of integrase in mammalian cells; FIG. 4C depicts plasmid pBCPB+, an intramolecular integration assay vector; FIG. 4D shows plasmid p220KattBfull, an EBV vector bearing attB, the target for integration events; FIG. 4E shows plasmid pTSAD, the donor for integration events, bearing attP. Kan^R, Amp^R, Chlor^R and Hyg^R are genes for resistance to kanamycin, ampicillin, chloramphenicol, and hygromycin, respectively.

FIG. 5 shows along the vertical axis the percent recombination obtained in the intramolecular integration assay in *E. coli,* described in Example 6, when various shortened versions of φC31 attB (left) and attP (right) were tested. The name of each site tested corresponds to the length of the att site in basepairs. The A and B of B33 indicate sites where the reduction of the site length from 34-bp to 33-bp occurred at the left or right ends of the site, respectively. Similar nomenclature is used for P39A and P39B. Full refers to the full length attB.

FIG. 6 shows the percent recombination obtained in the intramolecular integration assay performed in *E. coli* when various substitutions in the attB and/or attP cores were made. The first column shows the recombination frequency when attB bears the mutant sequence shown and attP remains wild-type, the second column shows the recombination frequency when attD bears the mutant sequence, while the third column shows the recombination frequency when both attB and attP bear the mutant core sequence shown. nd=not done. As the figure indicates, most changes in the core region are not well tolerated.

FIG. 7 shows the results of a bimolecular integration assay performed in human cells as described in the examples. Results are shown for human cells carrying three EBV plasmids, p220K, a negative control lacking attB; p220KattB35, which carries the minimally sized attB; and p220KattBfull, carrying the full-sized attB. Integration frequencies are shown for experiments when no DNA was transfected, when either the integrase expression plasmid pCMVInt or the attP-bearing plasmid pTSAD alone was transfected, or when both pCMVInt and pTSAD together were transfected. Only the latter conditions, in the presence of a plasmid bearing attB, lead to integration events. Integration frequencies were corrected for transfection frequency to give the accurate corrected integration frequencies in the last column. p220KattBfull produced the highest integration frequency at 7.5%.

FIGS. 8A through 8B show pseudo-loxP sequences identified by computer search, as described in the Examples. The core sequences are shown in boldface type.

DETAILED DESCRIPTION OF THE INVENTION

Throughout this application, various publications, patents, and published patent applications are referred to by an identifying citation. The disclosures of these publications, patents, and published patent specifications referenced in this application are hereby incorporated by reference into the present disclosure to more fully describe the state of the art to which this invention pertains.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, cell biology and recombinant DNA, which are within the skill of the art. See, e.g., Sambrook, Fritsch, and Maniatis, MOLECULAR CLONING: A LABORATORY MANUAL, 2nd edition (1989); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, (F. M. Ausubel et al. eds., 1987); the series METHODS IN ENZYMOLOGY (Academic Press, Inc.); PCR 2: A PRACTICAL APPROACH (M. J. McPherson, B. D. Hames and G. R. Taylor eds., 1995) and ANIMAL CELL CULTURE (R. I. Freshney. Ed., 1987).

All publications, patents and patent applications cited herein, whether supra or infra, are hereby incorporated by reference in their entirety.

As used in this specification and the appended claims, the singular forms "a," "an" and "the" include plural references unless the content clearly dictates otherwise. Thus, for example, reference to "an antigen" includes a mixture of two or more such agents.

Definitions

"Recombinase" as used herein refers to a group of enzymes that can facilitate site specific recombination between defined sites, where the sites are physically separated on a single DNA molecule or where the sites reside on separate DNA molecules. The DNA sequences of the defined recombination sites are not necessarily identical. Within this group are several subfamilies including "Integrase" (including, for example, Cre and λ integrase) and "Resolvase/Invertase" (including, for example, φC31 integrase, R4 integrase, and TP-901 integrase).

By "wild-type recombination site (RS/WT)" is meant a recombination site normally used by an integrase or recombinase. For example, λ is a temperate bacteriophage that infects *E. coli*. The phage has one attachment site for recombination (attP) and the *E. coli* bacterial genome has an attachment site for recombination (attB). Both of these sites are wild-type recombination sites for λ integrase. In the context of the present invention, wild-type recombination sites occur in the homologous phage/bacteria system. Accordingly, wild-type recombination sites can be derived from the homologous system and associated with heterologous sequences, for example, the Att_B site can be placed in other systems to act as a substrate for the integrase.

By "pseudo-recombination site (RS/P)" is meant a site at which recombinase can facilitate recombination even though the site may not have a sequence identical to the sequence of its wild-type recombination site. A pseudo-recombination site is typically found in an organism heterologous to the native phage/bacterial system. For example, a φC31 integrase and vector carrying a φC31 wild-type recombination site can be placed into a eucaryotic cell. The wild-type recombination sequence aligns itself with a sequence in the eucaryotic cell genome and the integrase facilitates a recombination event. When the sequence from the genomic site, in the eucaryotic cell, where the integration of the vector took place (via a recombination event between the wild-type recombination site in the vector and the genome) is examined, the sequence at the genomic site typically has some identity to but may not be identical with the wild-type bacterial genome recombination site. The recombination site in the eucaryotic cell is considered to be a pseudo-recombination site at least because the eucaryotic cell is heterologous to the normal phage/bacterial cell system. The size of the pseudo-recombination site can be determined through the use of a variety of methods including, but not limited to, (i) sequence alignment comparisons, (ii) secondary structural comparisons, (iii) deletion or point mutation analysis to find the functional limits of the pseudo-recombination site, and (iv) combinations of the foregoing. Pseudo-recombination sites typically occur naturally in the genomes of eucaryotic cells (i.e., the sites are native to the genome) and are functionally identified as described herein (e.g., see Examples).

By "hybrid-recombination site (RS/H)" as used herein refers to a recombination site constructed from portions of wild-type and/or pseudo-recombination sites. As an example, a wild-type recombination site may have a short, core region flanked by palindromes. In one embodiment of a "hybrid-recombination site" the short, core region sequence of the hybrid-recombination site matches a core sequence of a pseudo-recombination site and the palindromes of the hybrid-recombination site match the wild-type recombination site. In an alternative embodiment, the hybrid-recombination site may be comprised of flanking sites derived from a pseudo-recombination site and a core region derived from a wild-type recombination site. Other combinations of such hybrid-recombination sites will be evident to those having ordinary skill in the art, in view of the teachings of the present specification.

A recombination site "native" to the genome, as used herein, means a recombination site that occurs naturally in the genome of a cell (i.e., the sites are not introduced into the genome, for example, by recombinant means.)

By "nucleic acid construct" it is meant a nucleic acid sequence that has been constructed to comprise one or more functional units not found together in nature. Examples include circular, double-stranded, extrachromosomal DNA molecules (plasmids), cosmids (plasmids containing COS sequences from lambda phage), viral genomes comprising non-native nucleic acid sequences, and the like.

By "nucleic acid fragment of interest" it is meant any nucleic acid fragment that one wishes to insert into a genome. Suitable examples of nucleic acid fragments of interest include therapeutic genes, marker genes, control regions, trait-producing fragments, and the like.

"Therapeutic genes" are those nucleic acid sequences which encode molecules that provide some therapeutic benefit to the host, including proteins, functional RNAs (antisense, hammerhead ribozymes), and the like. One well known example is the cystic fibrosis transmembrane conductance regulator (CFTR) gene. The primary physiological defect in cystic fibrosis is the failure of electrogenic chloride ion secretion across the epithelia of many organs, including the lungs. One of the most dangerous aspects of the disorder is the cycle of recurrent airway infections which gradually destroy lung function resulting in premature death. Cystic fibrosis is caused by a variety of mutations in the CFTR gene. Since the problems arising in cystic fibrosis result from mutations in a single gene, the possibility exists that the introduction of a normal copy of the gene into the lung

epithelia could provide a treatment for the disease, or effect a cure if the gene transfer was permanent.

Other disorders resulting from mutations in a single gene (known as monogenic disorders) include alpha-l-antitrypsin deficiency, chromic granulomatous disease, familial hypercholesterolemia, Fanconi anemia, Gaucher disease, Hunter syndrome, ornithine transcarbamylase deficiency, purine nucleoside phosphorylase deficiency, severe combined immunodeficiency disease (SCID)-ADA, X-linked SCID, hemophilia, and the like.

Therapeutic benefit in other disorders may also result from the addition of a protein-encoding therapeutic nucleic acid. For example, addition of a nucleic acid encoding an immunomodulating protein such as interleukin-2 may be of therapeutic benefit for patients suffering from different types of cancer.

A nucleic acid fragment of interest may additionally be a "marker nucleic acid" or "marker polypeptide". Marker genes encode proteins which can be easily detected in transformed cells and are, therefore, useful in the study of those cells. Marker genes are being used in bone marrow transplantation studies, for example, to investigate the biology of marrow reconstitution and the mechanism of relapse in patients. Examples of suitable marker genes include beta—galactosidase, green or yellow fluorescent proteins, chloramphenicol acetyl transferase, luciferase, and the like.

A nucleic acid fragment of interest may additionally be a control region. The term "control region" or "control element" includes all nucleic acid components which are operably linked to a DNA fragment and involved in the expression of a protein or RNA therefrom. An operable linkage is a linkage in which the regulatory DNA fragments and the DNA sought to be expressed are connected in such a way as to permit coding sequence (the nucleic acids encoding the amino acid sequence of a protein) expression. The precise nature of the regulatory regions needed for coding sequence expression may vary from organism to organism, but will in general include a promoter region that, in prokaryotes, contains both the promoter (which directs the initiation of RNA transcription) as well as the DNA that, when transcribed into RNA, will signal synthesis initiation. Such regions will normally include those 5' noncoding sequences involved with initiation of transcription and translation, such as the enhancer, TATA box, capping sequence, CAAT sequence, and the like.

Under some circumstances, the native genome sought to be modified contains a functional coding sequence but lacks the ability to control the expression of the sequence. In such cases it would be of benefit to modify the genome by the insertion of control region(s). Such sequences include any sequence that functions to modulate replication, transcriptional or translational regulation, and the like. Examples include promoters, signal sequences, propeptide sequences, transcription terminators, polyadenylation sequences, enhancer sequences, attenuatory sequences, intron splice site sequences, and the like.

A nucleic acid fragment of interest may additionally be a trait-producing sequence, by which it is meant a sequence conferring some non-native trait upon the organism or cell in which the protein encoded by the trait-producing sequence is expressed. The term "non-native" when used in the context of a trait-producing sequence means that the trait produced is different than one would find in an unmodified organism which can mean that the organism produces high amounts of a natural substance in comparison to an unmodified organism, or produces a non-natural substance. For example, the genome of a crop plant, such as corn, can be modified to produce higher amounts of an essential amino acid, thus creating a plant of higher nutritional quality, or could be modified to produce proteins not normally produced in plants, such as antibodies. (See U.S. Pat. No. 5,202,422 (issued Apr. 13, 1993); U.S. Pat. No. 5,639,947 (Jun. 17, 1997).) Likewise, the genome of industrially important microorganisms can be modified to make them more useful such as by inserting new metabolic pathways with the aim of producing novel metabolites or improving both new and existing processes such as the production of antibiotics and industrial enzymes. Other useful traits include herbicide resistance, antibiotic resistance, disease resistance, resistance to adverse environmental conditions (e.g., temperature, pH, salt, drought), and the like.

Methods of transforming cells are well known in the art. By "transformed" it is meant a heritable alteration in a cell resulting from the uptake of foreign DNA. Suitable methods include viral infection, transfection, conjugation, protoplast fusion, electroporation, particle gun technology, calcium phosphate precipitation, direct microinjection, and the like. The choice of method is generally dependent on the type of cell being transformed and the circumstances under which the transformation is taking place (i.e. in vitro, ex viva, or in vivo). A general discussion of these methods can be found in Ausubel, et al, Short Protocols in Molecular Biology, 3rd ed., Wiley & Sons, 1995.

The terms "nucleic acid molecule" and "polynucleotide" are used interchangeably and refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof. Polynucleotides may have any three-dimensional structure, and may perform any function, known or unknown. Non-limiting examples of polynucleotides include a gene, a gene fragment, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers.

A polynucleotide is typically composed of a specific sequence of four nucleotide bases: adenine (A); cytosine (C); guanine (G); and thymine (T) (uracil (U) for thymine (T) when the polynucleotide is RNA). Thus, the term polynucleotide sequence is the alphabetical representation of a polynucleotide molecule. This alphabetical representation can be input into databases in a computer having a central processing unit and used for bioinformatics applications such as functional genomics and homology searching.

A "coding sequence" or a sequence which "encodes" a selected polypeptide, is a nucleic acid molecule which is transcribed (in the case of DNA) and translated (in the case of mRNA) into a polypeptide, for example, in vivo when placed under the control of appropriate regulatory sequences (or "control elements"). The boundaries of the coding sequence are typically determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxy) terminus. A coding sequence can include, but is not limited to, cDNA from viral, procaryotic or eucaryotic mRNA, genomic DNA sequences from viral or procaryotic DNA, and even synthetic DNA sequences. A transcription termination sequence may be located 3' to the coding sequence. Other "control elements" may also be associated with a coding sequence. A DNA sequence encoding a polypeptide can be optimized for expression in a selected cell by using the codons preferred by the selected cell to represent the DNA copy of the desired polypeptide coding sequence. "Encoded by" refers to a nucleic acid sequence

which codes for a polypeptide sequence, wherein the polypeptide sequence or a portion thereof contains an amino acid sequence of at least 3 to 5 amino acids, more preferably at least 8 to 10 amino acids, and even more preferably at least 15 to 20 amino acids from a polypeptide encoded by the nucleic acid sequence. Also encompassed are polypeptide sequences which are immunologically identifiable with a polypeptide encoded by the sequence.

"Operably linked" refers to an arrangement of elements wherein the components so described are configured so as to perform their usual function. Thus, a given promoter that is operably linked to a coding sequence (e.g., a reporter expression cassette) is capable of effecting the expression of the coding sequence when the proper enzymes are present. The promoter or other control elements need not be contiguous with the coding sequence, so long as they function to direct the expression thereof. For example, intervening untranslated yet transcribed sequences can be present between the promoter sequence and the coding sequence and the promoter sequence can still be considered "operably linked" to the coding sequence.

A "vector" is capable of transferring gene sequences to target cells. Typically, "vector construct," "expression vector," and "gene transfer vector," mean any nucleic acid construct capable of directing the expression of a gene of interest and which can transfer gene sequences to target cells. Thus, the term includes cloning, and expression vehicles, as well as integrating vectors.

An "expression cassette" comprises any nucleic acid construct capable of directing the expression of a gene/coding sequence of interest. Such cassettes can be constructed into a "vector," "vector construct," "expression vector," or "gene transfer vector," in order to transfer the expression cassette into target cells. Thus, the term includes cloning and expression vehicles, as well as viral vectors.

Techniques for determining nucleic acid and amino acid "sequence identity" also are known in the art. Typically, such techniques include determining the nucleotide sequence of the mRNA for a gene and/or determining the amino acid sequence encoded thereby, and comparing these sequences to a second nucleotide or amino acid sequence. In general, "identity" refers to an exact nucleotide-to-nucleotide or amino acid-to-amino acid correspondence of two polynucleotides or polypeptide sequences, respectively. Two or more sequences (polynucleotide or amino acid) can be compared by determining their "percent identity." The percent identity of two sequences, whether nucleic acid or amino acid sequences, is the number of exact matches between two aligned sequences divided by the length of the shorter sequences and multiplied by 100. An approximate alignment for nucleic acid sequences is provided by the local homology algorithm of Smith and Waterman, *Advances in Applied Mathematics* 2:482–489 (1981). This algorithm can be applied to amino acid sequences by using the scoring matrix developed by Dayhoff, *Atlas of Protein Sequences and Structure,* M. O. Dayhoff ed., 5 suppl. 3:353–358, National Biomedical Research Foundation, Washington, D.C., USA, and normalized by Gribskov, *Nucl. Acids Res.* 14(6):6745–6763 (1986). An exemplary implementation of this algorithm to determine percent identity of a sequence is provided by the Genetics Computer Group (Madison, Wis.) in the "BestFit" utility application. The default parameters for this method are described in the Wisconsin Sequence Analysis Package Program Manual, Version 8 (1995) (available from Genetics Computer Group, Madison, Wis.). A preferred method of establishing percent identity in the context of the present invention is to use the MPSRCH package of programs copyrighted by the University of Edinburgh, developed by John F. Collins and Shane S. Sturrok, and distributed by IntelliGenetics, Inc. (Mountain View, Calif.). From this suite of packages the Smith-Waterman algorithm can be employed where default parameters are used for the scoring table (for example, gap open penalty of 12, gap extension penalty of one, and a gap of six). From the data generated the "Match" value reflects "sequence identity." Other suitable programs for calculating the percent identity or similarity between sequences are generally known in the art, for example, another alignment program is BLAST, used with default parameters. For example, BLASTN and BLASTP can be used using the following default parameters: genetic code=standard; filter=none; strand=both; cutoff=60; expect=10; Matrix=BLOSUM62; Descriptions=50 sequences; sort by=HIGH SCORE; Databases=non-redundant, GenBank+EMBL+DDBJ+PDB+GenBank CDS translations+Swiss protein+Spupdate+PIR. Details of these programs can be found at the WebSite of NCBI/NLM.

Alternatively, homology can be determined by hybridization of polynucleotides under conditions that form stable duplexes between homologous regions, followed by digestion with single-stranded-specific nuclease(s), and size determination of the digested fragments. Two DNA, or two polypeptide sequences are "substantially homologous" to each other when the sequences exhibit at least about 80%–85%, preferably at least about 85%–90%, more preferably at least about 90%–95%, and most preferably at least about 95%–98% sequence identity over a defined length of the molecules, as determined using the methods above. As used herein, substantially homologous also refers to sequences showing complete identity to the specified DNA or polypeptide sequence. DNA sequences that are substantially homologous can be identified in a Southern hybridization experiment under, for example, stringent conditions, as defined for that particular system. Defining appropriate hybridization conditions is within the skill of the art. See, e.g., Sambrook et al., supra; *DNA Cloning,* supra; *Nucleic Acid Hybridization,* supra.

Two nucleic acid fragments are considered to "selectively hybridize" as described herein. The degree of sequence identity between two nucleic acid molecules affects the efficiency and strength of hybridization events between such molecules. A partially identical nucleic acid sequence will at least partially inhibit a completely identical sequence from hybridizing to a target molecule. Inhibition of hybridization of the completely identical sequence can be assessed using hybridization assays that are well known in the art (e.g., Southern blot, Northern blot, solution hybridization, or the like, see Sambrook, et al., *Molecular Cloning: A Laboratory Manual,* Second Edition, (1989) Cold Spring Harbor, N.Y.). Such assays can be conducted using varying degrees of selectivity, for example, using conditions varying from low to high stringency. If conditions of low stringency are employed, the absence of non-specific binding can be assessed using a secondary probe that lacks even a partial degree of sequence identity (for example, a probe having less than about 30% sequence identity with the target molecule), such that, in the absence of non-specific binding events, the secondary probe will not hybridize to the target.

When utilizing a hybridization-based detection system, a nucleic acid probe is chosen that is complementary to a target nucleic acid sequence, and then by selection of appropriate conditions the probe and the target sequence "selectively hybridize," or bind, to each other to form a hybrid molecule. A nucleic acid molecule that is capable of

hybridizing selectively to a target sequence under "moderately stringent" typically hybridizes under conditions that allow detection of a target nucleic acid sequence of at least about 10–14 nucleotides in length having at least approximately 70% sequence identity with the sequence of the selected nucleic acid probe. Stringent hybridization conditions typically allow detection of target nucleic acid sequences of at least about 10–14 nucleotides in length having a sequence identity of greater than about 90–95% with the sequence of the selected nucleic acid probe. Hybridization conditions useful for probe/target hybridization where the probe and target have a specific degree of sequence identity, can be determined as is known in the art (see, for example, *Nucleic Acid Hybridization: A Practical Approach,* editors B. D. Hames and S. J. Higgins, (1985) Oxford; Washington, D.C.; IRL Press).

With respect to stringency conditions for hybridization, it is well known in the art that numerous equivalent conditions can be employed to establish a particular stringency by varying, for example, the following factors: the length and nature of probe and target sequences, base composition of the various sequences, concentrations of salts and other hybridization solution components, the presence or absence of blocking agents in the hybridization solutions (e.g., formamide, dextran sulfate, and polyethylene glycol), hybridization reaction temperature and time parameters, as well as, varying wash conditions. The selection of a particular set of hybridization conditions is selected following standard methods in the art (see, for example, Sambrook, et al., *Molecular Cloning: A Laboratory Manual,* Second Edition, (1989) Cold Spring Harbor, N.Y.)

A first polynucleotide is "derived from" second polynucleotide if it has the same or substantially the same basepair sequence as a region of the second polynucleotide, its cDNA, complements thereof, or if it displays sequence identity as described above.

A first polypeptide is "derived from" a second polypeptide if it is (i) encoded by a first polynucleotide derived from a second polynucleotide, or (ii) displays sequence identity to the second polypeptides as described above. In the present invention, when a recombinase is "derived from a phage" the recombinase need not be explicitly produced by the phage itself, the phage is simply considered to be the original source of the recombinase and coding sequences thereof. Recombinases can, for example, be produced recombinantly or synthetically, by methods known in the art, or alternatively, recombinases may be purified from phage infected bacterial cultures.

"Substantially purified" general refers to isolation of a substance (compound, polynucleotide, protein, polypeptide, polypeptide composition) such that the substance comprises the majority percent of the sample in which it resides. Typically in a sample a substantially purified component comprises 50%, preferably 80%–85%, more preferably 90–95% of the sample. Techniques for purifying polynucleotides and polypeptides of interest are well-known in the art and include, for example, ion-exchange chromatography, affinity chromatography and sedimentation according to density.

1.0.0 The Invention

The invention disclosed herein comprises a method of specifically modifying a genome. In one embodiment of the method, a cell having a target recombination sequence (designated attT) is transformed with a nucleic acid construct (a "targeting construct") comprising a second recombination sequence (designated attD) and one or more polynucleotides of interest. Into the same cell a recombinase is introduced that specifically recognizes the recombination sequences under conditions such that the nucleic acid sequence of interest is inserted into the genome via a recombination event between attT and attD. Alternatively, the recombinase can be introduced into the cell prior to or concurrent with introduction of the targeting construct transformation with the nucleic acid construct.

The method of the invention is based, in part, on the discovery that there exist in various genomes specific nucleic acid sequences, herein called pseudo-recombination sequences, that may be distinct from wild-type recombination sequences and that can be recognized by a site-specific recombinase and used to promote the insertion of heterologous genes or polynucleotides into the genome. The inventors have identified such pseudo-recombination sequences in a variety of organisms, including mammals and plants.

1.1.0 Recombinases

Two major families of site-specific recombinases from bacteria and unicellular yeasts have been described: the integrase family includes Cre, Flp, R, and λ integrase (Argos, et al., EMBO J. 5:433–440, 1986) and the resolvase/invertase family includes some phage integrases, such as, those of phages φC31, R4, and TP-901 (Hallet and Sherratt, FEMS Microbiol. Rev. 21:157–178, 1997). While not wishing to be bound by descriptions of mechanisms, strand exchange catalyzed by site specific recombinases typically occurs in two steps of (1) cleavage and (2) rejoining involving a covalent protein-DNA intermediate formed between the recombinase enzyme and the DNA strand(s).

The nature of the catalytic amino acid residue of the recombinase enzyme and the line of entry of the nucleophile can be different for the two recombinase families. For cleavage catalyzed by the invertase/resolvase family, for example, the nucleophile hydroxyl is derived from a serine and the leaving group is the 3'-OH of the deoxyribose. For the integrase family, the catalytic residue is, for example, a tyrosine and the leaving group is the 5'-OH. In both recombinase families, the rejoining step is the reverse of the cleavage step. Recombinases particularly useful in the practice of the invention are those that function in a wide variety of cell types, in part because they do not require any host specific factors. Suitable recombinases include Cre, Flp, R, and the integrases of phages φC31, TP901-1, R4, and the like. Some characteristics of the two recombinase families are discussed below.

1.1.1 Cre-like Recombinases

The recombinase activity of Cre has been studied as a model system for the integrases. Cre is a 38 kD protein isolated from bacteriophage P1. It catalyzes recombination at a 34 basepair stretch of DNA called loxP. The loxP site has the sequence 5'-ATAACTTCGTATA GCATACAT TATACGAAGTTAT-3' (SEQ ID NO:1) consisting of two thirteen basepair palindromic repeats flanking an eight basepair core sequence. The repeat sequences act as Cre binding sites with the crossover point occurring in the core. Each repeat appears to bind one protein molecule wherein the DNA substrate (one strand) is cleaved and a protein DNA intermediate is formed having a 3'-phosphotyrosine linkage between Cre and the cleaved DNA strand. Crystallography and other studies suggest that four proteins and two loxP sites form a synapsed structure in which the DNA resembles models of four-way Holliday-junction intermediates, followed by the exchange of a second set of strands to resolve the intermediate into recombinant products (see, Guo, et al, Nature 389:40–46, 1997). The asymmetry of the core region is responsible for directionality of the recombination reaction. If the two recombination sites are repeated in the same

orientation, the outcome of strand exchange is integration or excision. If the two sites are placed in the opposite orientation, the outcome is inversion of the sequence between the two sites (Yang and Mizuuchi, Structure 5:1401–1406, 1997).

Cre has been shown to be active in a wide variety of cellular backgrounds including yeast (Sauer, Mol. Cell. Biol. 7:2087–2096, 1987), plants (Albert, et al, Plant J. 7:649–659, 1995; Dale and Ow, Gene 91:79–85, 1990; Odell, et al, Mol. Gen. Genet. 223:369–378, 1990) and mammals, including both rodent and human cells (van Deursen, et al, Proc. Natl. Acad. Sci. USA 92:7376–7380, 1995; Agah, et al, J. Clin. Invest. 100:169–179, 1997; Baubonis, and Sauer, 21:2025–2029, 1993; Sauer and Henderson, New Biologist 2:441–449, 1990). As the loxP site is known only to occur in the P1 phage genome, use of the enzyme in other cell types requires the prior insertion of a loxP site into the genome, which using currently available technologies is generally a low-frequency and random event with all of the drawbacks inherent in such a procedure. The loxP site can be targeted to a specific location by using homologous recombination, but, again, that process occurs at a very low frequency.

Several studies have suggested the possibility that an exact match of the loxP sequence is not required for Cre-mediated recombination (Sternberg, et al, J. Mol. Biol. 150:487–507, 1981; Sauer, J. Mol. Biol. 223:911–928, 1992; Sauer, Nucleic Acids Research 24:4608–4613, 1996). The efficiency of recombination, however, has generally been three to four orders of magnitude less efficient than wild-type loxP. Sauer attempted to identify sequences similar to loxP in the human genome without success (Sauer, Nucleic Acids Research 24:4608–4613, 1996).

Flp, a recombinase of the integrase family with similar properties to Cre has been identified in strains of *Saccharomyces cerevisiae* that contain 2μ-circle DNA. Flp recognizes a DNA sequence consisting of two thirteen basepair inverted repeats flanking an eight basepair core sequence (5'-GAAGTTCCTATAC TTCTAGAA GAATAGGAACTTC-3' (SEQ ID NO:2) called FRT. A third repeat follows at the 3' end in the natural sequence but does not appear to be required for recombinase activity. Like Cre, Flp is functional in a wide variety of systems including bacteria (Huang, et al, J Bacteriology 179:6076–6083, 1997), insects (Golic and Lindquist, Cell 59:499–509, 1989; Golic and Golic, Genetics 144:1693–1711, 1996), plants (Lyznik, et al, Nucleic Acids Res 21:969–975, 1993) and mammals. These studies have likewise required that a FRT sequence be inserted into the genome to be modified.

A related recombinase, known as R, is encoded by the pSR1 plasmid of the yeast *Zygosaccharomyces rouxii* (Araki, et al., J. Mol. Biol. 182:191–203, 1985, herein incorporated by reference). This recombinase may have properties similar to those described above.

In the context of the present invention, when a recombinase normally facilitates recombination between two recombination sites and the sites are essentially the same (e.g., loxP and Cre), the sites are designated recombinase-mediated-recombination sites (RMRS).

1.1.2 Resolvase/Integrase Recombinases

Unlike the Cre/λ integrase family of recombinases, members of the resolvase subfamily of recombinase enzymes typically contain an N-terminal catalytic domain having a high degree (>35%) of sequence homology among the subfamily members (Crellin and Rood, J Bacteriology 179 (16):5148–5156, 1997; Christiansen, et al, J. Bacteriology 178(17):5164–5S173, 1996). Like some of the Cre-type recombinases, however, some resolvases do not require host specific accessory factors (Thorpe and Smith, PNAS USA 95:5505–5510, 1998).

The process of strand exchange used by the resolvases is somewhat different than the process used by Cre. This process is described but is not intended to be limiting. The resolvases usually make cuts close to the center of the crossover site, and the top and bottom strand cuts are often staggered by 2 basepairs, leaving recessed 5' ends. A protein-DNA linkage is formed between phosphodiester from the 5' DNA end and a conserved serine residue close to the amino terminus of the recombinase. As with the Cre-like invertases, two protein units are bound at each crossover site, however, no equivalent to the Holiday junction intermediate is formed (see Stark, et al, Trends in Genetics 8(12):432–439, 1992, incorporated by reference herein).

The nucleic acid sequences recognized as recombination sites by a subset of the resolvase family, including some phage integrases, differ in several ways from the recombination site recognized by Cre. The sites used for recognition and recombination of the phage and bacterial DNAs (the native host system) are generally non-identical, although they typically have a common core region of nucleic acids. The bacterial sequence is generally called the attB sequence (bacterial attachment) and the phage sequence is called the attP sequence (phage attachment). Because they are different sequences, recombination will result in a stretch of nucleic acids (called attL or attR for left and right) that is neither an attB sequence or an attP sequence, and is probably functionally unrecognizable as a recombination site to the relevant enzyme, thus removing the possibility that the enzyme will catalyze a second recombination reaction that would reverse the first.

The individual resolvases and the nucleic acid sequences that they recognize have been less well characterized than Cre and Flp, although many of the core sequences have been identified. The core sequences of some of the resolvases useful in the practice of the invention can include, without limitation, the following sequences: ϕC31-5'-TTG; TP901-1-5'-TCAAT; and R4-5'-GAAGCAGTGGTA. (SEQ ID NO:3) (See Rausch and Lehmann, NAR 19:5187–5189, 1991; Shirai, et al, J Bacteriology 173(13):4237–4239, 1991; Crellin and Rood, J Bacteriology 179:5148–5156, 1997; Christiansen, et al, J. Bacteriology 176:1069–1076, 1994; Brondsted and Hammer, Applied & Environmental Microbiology 65:752–758, 1999; all of which are incorporated by reference herein.)

Several authors have suggested that integrase or resolvase (for example, ϕC31 integrase) can be used to modify bacterial genomes, such as, those of *E. coli* and actinomycetes (Mascarenhas and Olson, U.S. Pat. No. 5,470,727; Cox, et al, U.S. Pat. No. 5,190,871). However, there has been no suggestion that these enzymes would be useful in the modification of non-bacterial genomes.

1.1.3 Recombination Sites

The inventors have discovered native recombination sites existing in the genomes of a variety of organisms, where the native recombination site does not necessarily have a nucleotide sequence identical to the wild-type recombination sequences (for a given recombinase); but such native recombination sites are nonetheless sufficient to promote recombination meditated by the recombinase. Such recombination site sequences are referred to herein as "pseudo-recombination sequences." For a given recombinase, a pseudo-recombination sequence is functionally equivalent to a wild-type recombination sequence, occurs in an organism other than that in which the recombinase is found in

nature, and may have sequence variation relative to the wild type recombination sequences.

In the practice of the present invention, wild-type recombination sites, pseudo-recombination sites, and hybrid-recombination sites can be used in a variety of ways in the construction of targeting vectors. Following here are non-limiting examples of how these sites may be employed in the practice of the present invention.

Identification of pseudo-recombination sequences can be accomplished, for example, by using sequence alignment and analysis, where the query sequence is the recombination site of interest (for example, a recombinase-mediated-recombination site (RMRS; e.g., loxP), or either attB and/or attP of a phage/bacterial system). Following here are some examples: if a genomic recombination site (generally designated attT) is identified using attB, then that attT site is said to be a pseudo-attB site; if a genomic recombination site is identified using attP, then that attT site is said to be a pseudo-attP site; and, if a genomic recombination site is identified using an RMRS (e.g., loxP), then that attT site is said to be a pseudo-RMRS site (e.g., pseudo-loxP).

In one aspect of the present invention, the recombinase (for example, Cre) recognizes a recombination site having the following structure: flanking sequence palindrome—core sequence—flanking sequence palindrome. Such recombination sites typically comprise two approximately 10–20 base pair stretches having some palindromic character which flank an approximately 3–15 base pair core sequence.

In this aspect of the present invention, the genome of a target cell is searched for sequences having sequence identity to the selected recombination site for a given recombinase, for example, loxP (Example 1; FIG. 8). The cellular target recombination site (attT: in this example, a pseudo-loxP site) accordingly has a defined sequence. To practice the genome modification method of the present invention, a recombination sequence is placed in the targeting vector. This recombination sequence, attD, can take many forms but must be capable of participating in site specific recombination with the genomic site (attT) where the recombination is mediated by the appropriate recombinase. In this regard, non-limiting examples of attD sites include, but are not limited to, the following: attD core sequence matches the pseudo-recombination site core sequence, flanking sequences in the targeting construct are wild-type recombination sequences (this construct represents a hybrid-recombination site); or, attD core sequence matches the pseudo-recombination site core sequence, flanking sequences in the targeting construct match the pseudo-recombination site flanking sequences. Further, the core sequences between attT and attD are generally essentially the same and the flanking sequences for attD may be combinations of flanking sequences from wild-type and pseudo-recombination site sources.

The recombinase-mediated-recombination site (RMRS) of this type of recombinase, for example, Cre and Cre-like recombinases, can have the following structure: a first DNA sequence (RMRS5'), a core region A, and a second DNA sequence (RMRS3') in the relative order RMRS5'-core region A-RMRS3'. Such recombination sites typically comprise two approximately 10–20 base pair regions having palindromic characteristics (e.g., RMRS5' and RMRS3') which flank an approximately 3–15 basepair core sequence (for example, core region A). In one embodiment, e.g., when employing Cre, hybrid-recombination sites may be used where the palindromic sequences are derived from a wild-type recombination site and the core sequence is derived from a pseudo-recombination site.

Without being bound to any particular theory or mechanism of action, when such a nucleic acid construct is provided to a cell along with a site-specific recombinase, it is possible that the recombinase recognizes and binds to the flanking sequences of both hybrid-recombination sequence and the pseudo-recombination sequence from which the basepair core sequence was derived, and catalyzes the recombination between the two.

In one embodiment the attD (in the targeting construct) is a hybrid-lox sequence comprising two wild-type thirteen basepair loxP palindromes flanking a heterologous core sequence, where the core sequence corresponds to the core sequence of the pseudo-recombination sequence of attT (in the cell target). In a second embodiment the attD (in the targeting construct) is a hybrid-FRT sequence comprising two or three wild-type thirteen basepair palindromes flanking a heterologous core sequence, where the core sequences correspond to the core sequence of the pseudo-recombination sequence of attT (in the cell target).

Example 2 describes methods for testing whether a putative recombination site is functional as a pseudo-recombination site for recombination mediated by the selected site specific recombinase and also methods for assessing the efficiency of recombination.

In a second aspect of the present invention, the recombinase (for example, ϕC31) recognizes a recombination site where sequence of the 5' region of the recombination site can differ from the sequence of the 3' region of the recombination sequence. For example, for the phage ϕC31 attP (the phage attachment site), the core region is 5'-TTG-3' the flanking sequences on either side are represented here as attP5' and attP3', the structure of the attP recombination site is, accordingly, attP5'-TTG-attP3'. Correspondingly, for the native bacterial genomic target site (attB) the core region is 5'-TTG-3', and the flanking sequences on either side are represented here as attB5' and attB3', the structure of the attB recombination site is, accordingly, attB5'-TTG-attB3'. After a single-site, ϕC31 integrase mediated, recombination event takes place the result is the following recombination product: attB5'-TTG-attP3'{ϕC31 vector sequences}attP5'-TTG-attB3'. Typically, after recombination the post-recombination recombination sites are no longer able to act as substrate for the ϕC31 recombinase. This results in stable integration with little or no recombinase mediated excision. These structures are represented in a more generic way as follows: circular targeting vector comprising the recombination site (attD) and a polynucleotide of interest—attD5'-core-attD3'; pseudo-recombination site (attT)—attT5'-core-attT3'; post recombination structure—attT5'-recombination product site (e.g., core)-attD3'{polynucleotide sequences of interest}attD5'-recombination product site (e.g., core)-attT3'. The recombination product site sequence can comprise a core identical to the original core sequence. However, the complete post-recombination, recombination sites(for example, attT5'-recombination product site (e.g., core)-attD3') generally no longer provide a usable substrate for the recombinase.

In this aspect, when selecting pseudo-recombination sites in a target cell (attT), the genomic sequences of the target cell can be searched for suitable pseudo-recombination sites using either the attP or attB sequences associated with a particular recombinase. Functional sizes and the amount of heterogeneity that can be tolerated in these recombination sequences can be evaluated, for example, as described in Examples 8 and 9.

When a pseudo-recombination site is identified using either attP or attB search sequences, the other recombination

site can be used in the targeting construct. For example, if attP for a selected recombinase is used to identify a pseudo-recombination site in the target cell genome, then the wild-type attB sequence can be used in the targeting construct. In an alternative example, if attB for a selected recombinase is used to identify a pseudo-recombination site in the target cell genome, then the wild-type attP sequence can be used in the targeting construct.

The targeting constructs contemplated by the invention may contain additional nucleic acid fragments such as control sequences, marker sequences, selection sequences and the like as discussed below.

1.2.0 Targeting Constructs and Methods of the Present Invention

The present invention also provides means for targeted insertion of a polynucleotide (or nucleic acid sequence(s)) of interest into a genome by, for example, (i) providing a recombinase, wherein the recombinase is capable of facilitating recombination between a first recombination site and a second recombination site, (ii) providing a targeting construct having a first recombination sequence and a polynucleotide of interest, (iii) introducing the recombinase and the targeting construct into a cell which contains in its nucleic acid the second recombination site, wherein said introducing is done under conditions that allow the recombinase to facilitate a recombination event between the first and second recombination sites.

Historically, the attachment site in a bacterial genome is designated "attB" and in a corresponding bacteriophage the site is designated "attP". A recombination site in a cell of interest is designated herein as "attT". A recombination site in a targeting vector is referred to herein as "attD".

In one aspect of the present invention, at least one pseudo-recombination site for a selected recombinase is identified in a target cell of interest (attT). These sites can be identified by several methods including searching all known sequences derived from the cell of interest against a wild-type recombination site (e.g., attB or attP) for a selected recombinase (e.g., as described in Example 1). The functionality of pseudo-recombination sites identified in this way can then be empirically evaluated following the teachings of the present specification to determine their ability to participate in a recombinase-mediated recombination event.

1.2.1 Targeting Constructs of the Present Invention

A targeting construct, to direct integration to this pseudo-recombination site, would then comprise a recombination site (attD) wherein the recombinase can facilitate a recombination event between attT and attD, and a polynucleotide of interest. Polynucleotides of interest can include, but are not limited to, expression cassettes encoding polypeptide products. The targeting constructs are typically circular and may also contain selectable markers, an origin of replication, and other elements. Targeting constructs of the present invention are typically circular.

A variety of expression vectors are suitable for use in the practice of the present invention, both for prokaryotic expression and eukaryotic expression. In general, the targeting construct will have one or more of the following features: a promoter, promoter-enhancer sequences, a selection marker sequence, an origin of replication, an inducible element sequence, an epitope—tag sequence, and the like.

Promoter and promoter-enhancer sequences are DNA sequences to which RNA polymerase binds and initiates transcription. The promoter determines the polarity of the transcript by specifying which strand will be transcribed. Bacterial promoters consist of consensus sequences, −35 and −10 nucleotides relative to the transcriptional start, which are bound by a specific sigma factor and RNA polymerase. Eukaryotic promoters are more complex. Most promoters utilized in expression vectors are transcribed by RNA polymerase II. General transcription factors (GTFS) first bind specific sequences near the start and then recruit the binding of RNA polymerase II. In addition to these minimal promoter elements, small sequence elements are recognized specifically by modular DNA-binding/trans-activating proteins (e.g. AP-1, SP-1) that regulate the activity of a given promoter. Viral promoters serve the same function as bacterial or eukaryotic promoters and either provide a specific RNA polymerase in trans (bacteriophage T7) or recruit cellular factors and RNA polymerase (SV40, RSV, CMV). Viral promoters may be preferred as they are generally particularly strong promoters.

Promoters may be, furthermore, either constitutive or regulatable (i.e., inducible or derepressible). Inducible elements are DNA sequence elements which act in conjunction with promoters and bind either repressors (e.g. lacO/LAC Iq repressor system in *E. coli*) or inducers (e.g. gal1/GAL4 inducer system in yeast). In either case, transcription is virtually "shut off" until the promoter is derepressed or induced, at which point transcription is "turned-on."

Examples of constitutive promoters include the int promoter of bacteriophage λ, the bla promoter of the β-lactamase gene sequence of pBR322, the CAT promoter of the chloramphenicol acetyl transferase gene sequence of pPR325, and the like. Examples of inducible prokaryotic promoters include the major right and left promoters of bacteriophage (P_L and P_R), the trp, recA, lacZ, AraC and gal promoters of *E. coli*, the α-amylase (Ulmanen Elt at., J. Bacteriol. 162:176–182, 1985) and the sigma-28-specific promoters of *B. subtilis* (Gilman et al., Gene sequence 32:11–20(1984)), the promoters of the bacteriophages of Bacillus (Gryczan, In: The Molecular Biology of the Bacilli, Academic Press, Inc., NY (1982)), Streptomyces promoters (Ward et at., Mol. Gen. Genet. 203:468–478, 1986), and the like. Exemplary prokaryotic promoters are reviewed by Glick (J. Ind. Microtiot. 1:277–282, 1987); Cenatiempo (Biochimie 68:505–516, 1986); and Gottesman (Ann. Rev. Genet. 18:415–442, 1984).

Preferred eukaryotic promoters include, but are not limited to, the following: the promoter of the mouse metallothionein I gene sequence (Hamer et al., J. Mol. Appl. Gen. 1:273–288, 1982); the TK promoter of Herpes virus (McKnight, Cell 31:355–365, 1982); the SV40 early promoter (Benoist et al., Nature (London) 290:304–310, 1981); the yeast gall gene sequence promoter (Johnston et al., Proc. Natl. Acad. Sci. (USA) 79:6971–6975, 1982); Silver et al., Proc. Natl. Acad. Sci. (USA) 81:5951–59SS, 1984), the CMV promoter, the EF-1 promoter, Ecdysone-responsive promoter(s), tetracycline-responsive promoter, and the like.

Exemplary promoters for use in the present invention are selected such that they are functional in cell type (and/or animal or plant) into which they are being introduced.

Selection markers are valuable elements in expression vectors as they provide a means to select for growth of only those cells that contain a vector. Such markers are of two types: drug resistance and auxotrophic. A drug resistance marker enables cells to detoxify an exogenously added drug that would otherwise kill the cell. Auxotrophic markers allow cells to synthesize an essential component (usually an amino acid) while grown in media that lacks that essential component.

Common selectable marker genes include those for resistance to antibiotics such as ampicillin, tetracycline, kanamycin, bleomycin, streptomycin, hygromycin,

neomycin, Zeocin™, and the like. Selectable auxotrophic genes include, for example, hisD, that allows growth in histidine free media in the presence of histidinol.

A further element useful in an expression vector is an origin of replication. Replication origins are unique DNA segments that contain multiple short repeated sequences that are recognized by multimeric origin-binding proteins and that play a key role in assembling DNA replication enzymes at the origin site. Suitable origins of replication for use in expression vectors employed herein include *E. coli* oriC, colE1 plasmid origin, 2μ and ARS (both useful in yeast systems), sf1, SV40, EBV oriP (useful in mammalian systems), and the like.

Epitope tags are short peptide sequences that are recognized by epitope specific antibodies. A fusion protein comprising a recombinant protein and an epitope tag can be simply and easily purified using an antibody bound to a chromatography resin. The presence of the epitope tag furthermore allows the recombinant protein to be detected in subsequent assays, such as Western blots, without having to produce an antibody specific for the recombinant protein itself. Examples of commonly used epitope tags include V5, glutathione-S-transferase (GST), hemaglutinin (HA), the peptide Phe-His-His-Thr-Thr, chitin binding domain, and the like.

A further useful element in an expression vector is a multiple cloning site or polylinker. Synthetic DNA encoding a series of restriction endonuclease recognition sites is inserted into a plasmid vector, for example, downstream of the promoter element. These sites are engineered for convenient cloning of DNA into the vector at a specific position.

The foregoing elements can be combined to produce expression vectors suitable for use in the methods of the invention. Those of skill in the art would be able to select and combine the elements suitable for use in their particular system in view of the teachings of the present specification. Suitable prokaryotic vectors include plasmids such as those capable of replication in *E. coli* (for example, pBR322, ColE1, pSC101, PACYC 184, itVX, PRSET, pBAD (Invitrogen, Carlsbad, Calif.) and the like). Such plasmids are disclosed by Sambrook (cf. "Molecular Cloning: A Laboratory Manual," second edition, edited by Sambrook, Fritsch, & Maniatis, Cold Spring Harbor Laboratory, (1989)). Bacillus plasmids include pCl94, pC221, pTI27, and the like, and are disclosed by Gryczan (In: The Molecular Biology of the Bacilli, Academic Press, NY (1982), pp. 307–329). Suitable Streptomyces plasmids include pIi10l (Kendall et al., J. Bacteriol. 169:4177–4183, 1987), and streptomyces bacteriophages such as φC31 (Chater et al., In: Sixth International Symposium on Actinomycetales Biology, Akademiai Kaido, Budapest, Hungary (1986), pp. 45–54). Pseudomonas plasmids are reviewed by John et al. (Rev. Infect. Dis. 8:693–704, 1986), and Izaki (Jpn. J. Bacteriol. 33:729–742, 1978).

Suitable eukaryotic plasmids include, for example, BPV, EBV, vaccinia, SV40, 2-micron circle, pcDNA3.1, pcDNA3.1/GS, pYES2/GS, pMT, p IND, pIND(Spl), pVgRXR (Invitrogen), and the like, or their derivatives. Such plasmids are well known in the art (Botstein et al., Miami Wntr. SyTnp. 19:265–274, 1982; Broach, In: "The Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance", Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., p. 445–470, 1981; Broach, Cell 28:203–204, 1982; Dilon et al., J. Clin. Hematol. Oncol.10:39–48, 1980; Maniatis, In: Cell Biology: A Comprehensive Treatise, Vol. 3, Gene Sequence Expression, Academic Press, NY, pp. 563–608,1980.

The targeting cassettes described herein can be constructed utilizing methodologies known in the art of molecular biology (see, for example, Ausubel or Maniatis) in view of the teachings of the specification. As described above, the targeting constructs are assembled by inserting, into a suitable vector backbone, an attD (recombination site), polynucleotides encoding sequences of interest operably linked to a promoter of interest; and, optionally a sequence encoding a positive selection marker.

A preferred method of obtaining polynucleotides, including suitable regulatory sequences (e.g., promoters) is PCR. General procedures for PCR are taught in MacPherson et al., PCR: A PRACTICAL APPROACH, (IRL Press at Oxford University Press, (1991)). PCR conditions for each application reaction may be empirically determined. A number of parameters influence the success of a reaction. Among these parameters are annealing temperature and time, extension time, Mg2+ and ATP concentration, pH, and the relative concentration of primers, templates and deoxyribonucleotides. After amplification, the resulting fragments can be detected by agarose gel electrophoresis followed by visualization with ethidium bromide staining and ultraviolet illumination.

The expression cassettes, targeting constructs, vectors, recombinases and recombinase-coding sequences of the present invention can be formulated into kits. Components of such kits can include, but are not limited to, containers, instructions, solutions, buffers, disposables, and hardware.

1.2.2 Introducing Recombinases

In the methods of the invention a site-specific recombinase is introduced into a cell whose genome is to be modified. Methods of introducing functional proteins into cells are well known in the art. Introduction of purified recombinase protein ensures a transient presence of the protein and its function, which is often a preferred embodiment. Alternatively, a gene encoding the recombinase can be included in an expression vector used to transform the cell. It is generally preferred that the recombinase be present for only such time as is necessary for insertion of the nucleic acid fragments into the genome being modified. Thus, the lack of permanence associated with most expression vectors is not expected to be detrimental.

The recombinases used in the practice of the present invention can be introduced into a target cell before, concurrently with, or after the introduction of a targeting vector. The recombinase can be directly introduced into a cell as a protein, for example, using liposomes, coated particles, or microinjection. Alternately, a polynucleotide encoding the recombinase can be introduced into the cell using a suitable expression vector. The targeting vector components described above are useful in the construction of expression cassettes containing sequences encoding a recombinase of interest. Expression of the recombinase is typically desired to be transient. Accordingly, vectors providing transient expression of the recombinase are preferred in the practice of the present invention. However, expression of the recombinase can be regulated in other ways, for example, by placing the expression of the recombinase under the control of a regulatable promoter (i.e., a promoter whose expression can be selectively induced or repressed).

Sequences encoding recombinases useful in the practice of the present invention are known and include, but are not limited to, the following: Cre—Sternberg, et al., J. Mol. Biol. 187:197–212; φC31—Kuhstoss and Rao, J. Mol. Biol. 222:897–908, 1991; TP901-1—Christiansen, et al., J. Bact. 178:5164–5173, 1996; R4—Matsuura, et al., J. Bact. 178:3374–3376, 1996.

Recombinases for use in the practice of the present invention can be produced recombinantly or purified as previously described. Polypeptides having the desired recombinase activity can be purified to a desired degree of purity by methods known in the art of protein ammonium sulfate precipitation, purification, including, but not limited to, size fractionation, affinity chromatography, HPLC, ion exchange chromatography, heparin agarose affinity chromatography (e.g., Thorpe & Smith, Proc. Nat. Acad. Sci. 95:5505–5510, 1998.)

1.2.3 Cells

Cells suitable for modification employing the methods of the invention include both prokaryotic cells and eukaryotic cells, provided that the cell's genome contains a pseudo-recombination sequence. Prokaryotic cells are cells that lack a defined nucleus. Examples of suitable prokaryotic cells include bacterial cells, mycoplasmal cells and archaebacterial cells. Particularly preferred prokaryotic cells include those that are useful either in various types of test systems (discussed in greater detail below) or those that have some industrial utility such as *Klebsiella oxytoca* (ethanol production), *Clostridium acetobutylicum* (butanol production), and the like (see Green and Bennet, Biotech & Bioengineering 58:215–221, 1998; Ingram, et al, Biotech & Bioengineering 58:204–206, 1998). Suitable eukaryotic cells include both animal cells (such as from insect, rodent, cow, goat, rabbit, sheep, non-human primate, human, and the like) and plant cells (such as rice, corn, cotton, tobacco, tomato, potato, and the like). Cell types applicable to particular purposes are discussed in greater detail below

Yet another embodiment of the invention comprises isolated genetically engineered cells. Suitable cells may be prokaryotic or eukaryotic, as discussed above. The genetically engineered cells of the invention may be unicellular organisms or may be derived from multicellular organisms. By "isolated" in reference to genetically engineered cells derived from multicellular organisms it is meant the cells are outside a living body, whether plant or animal, and in an artificial environment. The use of the term isolated does not imply that the genetically engineered cells are the only cells present.

In one embodiment, the genetically engineered cells of the invention contain any one of the nucleic acid constructs of the invention. In a second embodiment, a recombinase that specifically recognizes recombination sequences is introduced into genetically engineered cells containing one of the nucleic acid constructs of the invention under conditions such that the nucleic acid sequence(s) of interest will be inserted into the genome. Thus, the genetically engineered cells possess a modified genome. Methods of introducing such a recombinase are well known in the art and are discussed above.

The genetically engineered cells of the invention can be employed in a variety of ways. Unicellular organisms can be modified to produce commercially valuable substances such as recombinant proteins, industrial solvents, industrially useful enzymes, and the like. Preferred unicellular organisms include fungi such as yeast (for example, *S. pombe, Pichia pastoris, S. cerevisiae* (such as INVSc1), and the like) Aspergillis, and the like, and bacteria such as Klebsiella, Streptomyces, and the like.

Isolated cells from multicellular organisms can be similarly useful, including insect cells, mammalian cells and plant cells. Mammalian cells that may be useful include those derived from rodents, primates and the like. They include HeLa cells, cells of fibroblast origin such as VERO, 3T3 or CHOK1, HEK 293 cells or cells of lymphoid origin (such as 32D cells) and their derivatives. Preferred mammalian host cells include nonadherent cells such as CHO, 32D, and the like.

In addition, plant cells are also available as hosts, and control sequences compatible with plant cells are available, such as the cauliflower mosaic virus 35S and 19S, nopaline synthase promoter and polyadenylation signal sequences, and the like. Appropriate transgenic plant cells can be used to produce transgenic plants.

Another preferred host is an insect cell, for example from the Drosophila larvae. Using insect cells as hosts, the Drosophila alcohol dehydrogenase promoter can be used (Rubin, Science 240:1453–1459, 1988). Alternatively, baculovirus vectors can be engineered to express large amounts of peptide encoded by a desired nucleic acid sequence in insect cells (Jasny, Science 238:1653, 1987); Miller et al., In: Genetic Engineering (1986), Setlow, J. K., et al., eds., Plenum, Vol. 8, pp. 277–297).

The genetically engineered cells of the invention are additionally useful as tools to screen for substances capable of modulating the activity of a protein encoded by a nucleic acid fragment of interest. Thus, an additional embodiment of the invention comprises methods of screening comprising contacting genetically engineered cells of the invention with a test substance and monitoring the cells for a change in cell phenotype, cell proliferation, cell differentiation, enzymatic activity of the protein or the interaction between the protein and a natural binding partner of the protein when compared to test cells not contacted with the test substance.

A variety of test substances can be evaluated using the genetically engineered cells of the invention including peptides, proteins, antibodies, low molecular weight organic compounds, natural products derived from, for example, fungal or plant cells, and the like. By "low molecular weight organic compound" it is, meant a chemical species with a molecular weight of generally less than 500–1000. Sources of test substances are well known to those of skill in the art.

Various assay methods employing cells are also well known by those skilled in the art. They include, for example, assays for enzymatic activity (Hirth, et al, U.S. Pat. No. 5,763,198, issued Jun. 9, 1998), assays for binding of a test substance to a protein expressed by the genetically engineered cells, assays for transcriptional activation of a reporter gene, and the like.

Cells modified by the methods of the present invention can be maintained under conditions that, for example, (i) keep them alive but do not promote growth, (ii) promote growth of the cells, and/or (iii) cause the cells to differentiate or dedifferentiate. Cell culture conditions are typically permissive for the action of the recombinase in the cells, although regulation of the activity of the recombinase may also be modulated by culture conditions (e.g., raising or lowering the temperature at which the cells are cultured). For a given cell, cell-type, tissue, or organism, culture conditions are known in the art.

2.0.0 Transgenic Plants and Non-Human Animals

In another embodiment, the present invention comprises transgenic plants and nonhuman transgenic animals whose genomes have been modified by employing the methods and compositions of the invention. Transgenic animals may be produced employing the methods of the present invention to serve as a model system for the study of various disorders and for screening of drugs that modulate such disorders.

A "transgenic" plant or animal refers to a genetically engineered plant or animal, or offspring of genetically engineered plants or animals. A transgenic plant or animal usually contains material from at least one unrelated

organism, such as, from a virus. The term "animal" as used in the context of transgenic organisms means all species except human. It also includes an individual animal in all stages of development, including embryonic and fetal stages. Farm animals (e.g., chickens, pigs, goats, sheep, cows, horses, rabbits and the like), rodents (such as mice), and domestic pets (e.g., cats and dogs) are included within the scope of the present invention. In a preferred embodiment, the animal is a mouse or a rat.

The term "chimeric" plant or animal is used to refer to plants or animals in which the heterologous gene is found, or in which the heterologous gene is expressed in some but not all cells of the plant or animal.

The term transgenic animal also includes a germ cell line transgenic animal. A "germ cell line transgenic animal" is a transgenic animal in which the genetic information provided by the invention method has been taken up and incorporated into a germ line cell, therefore conferring the ability to transfer the information to offspring. If such offspring, in fact, possess some or all of that information, then they, too, are transgenic animals.

Methods of generating transgenic plants and animals are known in the art and can be used in combination with the teachings of the present application.

In one embodiment, a transgenic animal of the present invention is produced by introducing into a single cell embryo a nucleic acid construct, comprising an attD recombination site capable of recombining with an attT recombination site found within the genome of the organism from which the cell was derived and a nucleic acid fragment of interest, in a manner such that the nucleic acid fragment of interest is stably integrated into the DNA of germ line cells of the mature animal and is inherited in normal Mendelian fashion. In this embodiment, the nucleic acid fragment of interest can be any one of the fragment described previously. Alternatively, the nucleic acid sequence of interest can encode an exogenous product that disrupts or interferes with expression of an endogenously produced protein of interest, yielding a transgenic animals with decreased expression of the protein of interest.

A variety of methods are available for the production of transgenic animals. A nucleic acid construct of the invention can be injected into the pronucleus, or cytoplasm, of a fertilized egg before fusion of the male and female pronuclei, or injected into the nucleus of an embryonic cell (e.g., the nucleus of a two-cell embryo) following the initiation of cell division (Brinster, et al., Proc. Nat. Acad. Sci. USA 82: 4438, 1985). Embryos can be infected with viruses, especially retroviruses, modified with an attD recombination site and a nucleic acid sequence of interest. The cell can further be treated with a site-specific recombinase as described above to promote integration of the nucleic acid sequence of interest into the genome.

By way of example only, to prepare a transgenic mouse, female mice are induced to superovulate. After being allowed to mate, the females are sacrificed by CO_2 asphyxiation or cervical dislocation and embryos are recovered from excised oviducts. Surrounding cumulus cells are removed. Pronuclear embryos are then washed and stored until the time of injection. Randomly cycling adult female mice are paired with vasectomized males. Recipient females are mated at the same time as donor females. Embryos then are transferred surgically. The procedure for generating transgenic rats is similar to that of mice. See Hammer, et al., Cell 63:1099–1112, 1990). Rodents suitable for transgenic experiments can be obtained from standard commercial sources such as Charles River (Wilmington, Mass.), Taconic (Germantown, N.Y.), Harlan Sprague Dawley (Indianapolis, Ind.), etc.

The procedures for manipulation of the rodent embryo and for microinjection of DNA into the pronucleus of the zygote are well known to those of ordinary skill in the art (Hogan, et al., supra). Microinjection procedures for fish, amphibian eggs and birds are detailed in Houdebine and Chourrout, Experientia 47:897–905, 1991). Other procedures for introduction of DNA into tissues of animals are described in U.S. Pat. No. 4,945,050 (Sandford et al., Jul. 30, 1990).

Totipotent or pluripotent stem cells derived from the inner cell mass of the embryo and stabilized in culture can be manipulated in culture to incorporate nucleic acid sequences employing invention methods. A transgenic animal can be produced from such cells through injection into a blastocyst that is then implanted into a foster mother and allowed to come to term.

Methods for the culturing of stem cells and the subsequent production of transgenic animals by the introduction of DNA into stem cells using methods such as electroporation, calcium phosphate/DNA precipitation, microinjection, liposome fusion, retroviral infection, and the like are also are well known to those of ordinary skill in the art. See, for example, Teratocarcinomas and Embryonic Stem Cells, A Practical Approach, E. J. Robertson, ed., IRL Press, 1987). Reviews of standard laboratory procedures for microinjection of heterologous DNAs into mammalian (mouse, pig, rabbit, sheep, goat, cow) fertilized ova include: Hogan et al., *Manipulating the Mouse Embryo* (Cold Spring Harbor Press 1986); Krimpenfort et al., 1991, Bio/Technology 9:86; Palmiter et al., 1985, Cell 41:343; Kraemer et al., *Genetic Manipulation of the Early Mammalian Embryo* (Cold Spring Harbor Laboratory Press 1985); Hammer et al., 1985, Nature, 315:680; Purcel et al., 1986, *Science*, 244:1281; Wagner et al., U.S. Pat. No. 5,175,385; Krimpenfort et al., U.S. Pat. No. 5,175,384, the respective contents of which are incorporated by reference.

The final phase of the procedure is to inject targeted ES cells into blastocysts and to transfer the blastocysts into pseudopregnant females. The resulting chimeric animals are bred and the offspring are analyzed by Southern blotting to identify individuals that carry the transgene. Procedures for the production of non-rodent mammals and other animals have been discussed by others (see Houdebine and Chourrout, supra; Pursel, et al., Science 244:1281–1288, 1989; and Simms, et al., Bio/Technology 6:179–183, 1988). Animals carrying the transgene can be identified by methods well known in the art, e.g., by dot blotting or Southern blotting.

The term transgenic as used herein additionally includes any organism whose genome has been altered by in vitro manipulation of the early embryo or fertilized egg or by any transgenic technology to induce a specific gene knockout. The term "gene knockout" as used herein, refers to the targeted disruption of a gene in vivo with loss of function that has been achieved by use of the invention vector. In one embodiment, transgenic animals having gene knockouts are those in which the target gene has been rendered nonfunctional by an insertion targeted to the gene to be rendered non-functional by targeting a pseudo-recombination site located within the gene sequence.

3.0.0 Gene Therapy and Disorders

A further embodiment of the invention comprises a method of treating a disorder in a subject in need of such treatment. In one embodiment of the method, at least one cell or cell type (or tissue, etc.) of the subject has a target recombination sequence (designated attT). This cell(s) is transformed with a nucleic acid construct (a "targeting

construct") comprising a second recombination sequence (designated attD) and one or more polynucleotides of interest (typically a therapeutic gene). Into the same cell a recombinase is introduced that specifically recognizes the recombination sequences under conditions such that the nucleic acid sequence of interest is inserted into the genome via a recombination event between attP and attD. Subjects treatable using the methods of the invention include both humans and non-human animals. Thus methods utilize the targeting constructs and recombinases of the present invention.

A variety of disorders may be treated by employing the method of the invention including monogenic disorders, infectious diseases, acquired disorders, cancer, and the like. Exemplary monogenic disorders include ADA deficiency, cystic fibrosis, familial-hypercholesterolemia, hemophilia, chronic ganulomatous disease, Duchenne muscular dystrophy, Fanconi anemia, sickle-cell anemia, Gaucher's disease, Hunter syndrome, X-linked SCID, and the like.

Infectious diseases treatable by employing the methods of the invention include infection with various types of virus including human T-cell lymphotropic virus, influenza virus, papilloma virus, hepatitis virus, herpes virus, Epstein-Bar virus, immunodeficiency viruses (HIV, and the like), cytomegalovirus, and the like. Also included are infections with other pathogenic organisms such as *Mycobacterium Tuberculosis, Mycoplasma pneumoniae,* and the like or parasites such as *Plasmadium falciparum,* and the like.

The term "acquired disorder" as used herein refers to a noncongenital disorder. Such disorders are generally considered more complex than monogenic disorders and may result from inappropriate or unwanted activity of one or more genes. Examples of such disorders include peripheral artery disease, rheumatoid arthritis, coronary artery disease, and the like.

A particular group of acquired disorders treatable by employing the methods of the invention include various cancers, including both solid tumors and hematopoietic cancers such as leukemias and lymphomas. Solid tumors that are treatable utilizing the invention method include carcinomas, sarcomas, osteomas, fibrosarcomas, chondrosarcomas, and the like. Specific cancers include breast cancer, brain cancer, lung cancer (non-small cell and small cell), colon cancer, pancreatic cancer, prostate cancer, gastric cancer, bladder cancer, kidney cancer, head and neck cancer, and the like.

The suitability of the particular place in the genome is dependent in part on the particular disorder being treated. For example, if the disorder is a monogenic disorder and the desired treatment is the addition of a therapeutic nucleic acid encoding a non-mutated form of the nucleic acid thought to be the causative agent of the disorder, a suitable place may be a region of the genome that does not encode any known protein and which allows for a reasonable expression level of the added nucleic acid. Methods of identifying suitable places in the genome are well known in the art and described further in the Examples below.

The nucleic acid construct useful in this embodiment is additionally comprised of one or more nucleic acid fragments of interest. Preferred nucleic acid fragments of interest for use in this embodiment are therapeutic genes and/or control regions, as previously defined. The choice of nucleic acid sequence will depend on the nature of the disorder to be treated. For example, a nucleic acid construct intended to treat hemophilia B, which is caused by a deficiency of coagulation factor IX, may comprise a nucleic acid fragment encoding functional factor IX. A nucleic acid construct intended to treat obstructive peripheral artery disease may comprise nucleic acid fragments encoding proteins that stimulate the growth of new blood vessels, such as, for example, vascular endothelial growth factor, platelet-derived growth factor, and the like. Those of skill in the art would readily recognize which nucleic acid fragments of interest would be useful in the treatment of a particular disorder.

The nucleic acid construct can be administered to the subject being treated using a variety of methods. Administration can take place in vivo or ex vivo. By "in vivo," it is meant in the living body of an animal. By "ex vivo" it is meant that cells or organs are modified outside of the body, such cells or organs are typically returned to a living body.

Methods for the therapeutic administration of nucleic acid constructs are well known in the art. Nucleic acid constructs can be delivered with cationic lipids (Goddard, et al, Gene Therapy, 4:1231–1236, 1997; Gorman, et al, Gene Therapy 4:983–992, 1997; Chadwick, et al, Gene Therapy 4:937–942, 1997; Gokhale, et al, Gene Therapy 4:1289–1299, 1997; Gao, and Huang, Gene Therapy 2:710–722, 1995, all of which are incorporated by reference herein), using viral vectors (Monahan, et al, Gene Therapy 4:40–49, 1997; Onodera, et al, Blood 91:30–36, 1998, all of which are incorporated by reference herein), by uptake of "naked DNA", and the like. Techniques well known in the art for the transfection of cells (see discussion above) can be used for the ex vivo administration of nucleic acid constructs. The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition. (See e.g. Fingl et al., 1975, in "The Pharmacological Basis of Therapeutics", Ch. 1 pl).

It should be noted that the attending physician would know how to and when to terminate, interrupt, or adjust administration due to toxicity, to organ dysfunction, and the like. Conversely, the attending physician would also know how to adjust treatment to higher levels if the clinical response were not adequate (precluding toxicity). The magnitude of an administered dose in the management of the disorder being treated will vary with the severity of the condition to be treated, with the route of administration, and the like. The severity of the condition may, for example, be evaluated, in part, by standard prognostic evaluation methods. Further, the dose and perhaps dose frequency will also vary according to the age, body weight, and response of the individual patient.

In general at least 1–10% of the cells targeted for genomic modification should be modified in the treatment of a disorder. Thus, the method and route of administration will optimally be chosen to modify at least 0.1–1% of the target cells per administration. In this way, the number of administrations can be held to a minimum in order to increase the efficiency and convenience of the treatment.

Depending on the specific conditions being treated, such agents may be formulated and administered systemically or locally. Techniques for formulation and administration may be found in "Remington's Pharmaceutical Sciences," 1990, 18th ed., Mack Publishing Co., Easton, Pa. Suitable routes may include oral, rectal, transdermal, vaginal, transmucosal, or intestinal administration; parenteral delivery, including intramuscular, subcutaneous, intramedullary injections, as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, intranasal, or intraocular injections, just to name a few.

The subject being treated will additionally be administered a recombinase that specifically recognizes the attP and attD recombination sequences that are selected for use. The

particular recombinase can be administered by including a nucleic acid encoding it as part of a nucleic acid construct, or as a protein to be taken up by the cells whose genome is to be modified. Methods and routes of administration will be similar to those described above for administration of a targeting construct comprising a recombination sequence and nucleic acid sequence of interest. The recombinase protein is likely to only be required for a limited period of time for integration of the nucleic acid sequence of interest. Therefore, if introduced as a recombinase gene, the vector carrying the recombinase gene will lack sequences mediating prolonged retention. For example, conventional plasmid DNA decays rapidly in most mammalian cells. The recombinase gene may also be equipped with gene expression sequences that limit its expression. For example, an inducible promoter can be used, so that recombinase expression can be temporally limited by limited exposure to the inducing agent. One such exemplary group of promoters are tetracycline-responsive promoters the expression of which can be regulated using tetracycline or doxycycline.

The invention will now be described in greater detail by reference to the following non-limiting Examples.

EXAMPLES

Example 1

Identification of Pseudo-recombination Sequences

The following example describes the identification of pseudo-loxP sequences by computer search. Similar procedures can be used to identify other pseudo-recombination sequences.

The findpatterns algorithm of the Wisconsin Software Package Version 9.0 developed by the Genetics Computer Group (GCG; Madison, Wis.), was used to screen all sequences in the GenBank database (Benson et al., 1998, Nucleic Acids Res. 26, 1–7). Default parameters are given below. Patterns resembling the wild-type loxP sequence, called pseudo-loxP sites (ψlox) herein, were sought. The results from two different search strategies (Patterns #1 and #2, see below) were pooled.

The wild-type loxP site is 34 base pairs long and consists of two identical thirteen-basepair palindromes, separated by an eight-basepair core. It has been demonstrated that, while strand cutting and exchange take place in the eight-basepair core, the DNA sequence of most of this core is not critical, as long as it matches between the two sites that are to recombine (Hoess et al., 1986, Nucleic Acids Res. 14, 2287–2300; Sauer, 1996, Nucleic Acids Res. 24, 4608–4613). Therefore, most of these bases were set as n's in the search algorithm. Nucleic acid constructs created using the principles embodied in the invention allow for full control over the sequence of the incoming lox site, as its eight-basepair core can be made to match that of the genomic site being targeted. This feature of the recombination reaction gives the desired level of specificity, allowing targeting of only one ψlox site in the genome.

Previous studies have suggested that the central bases of the thirteen-basepair palindrome, those closest to the eight-basepair core, are important for Cre recognition. Therefore, greater weight was given to matching the inner four or five positions of the palindrome.

Using search Pattern #1, a search was constructed in such a way that the sequences returned by the search program would only look for resemblance in the thirteen-basepair palindromic regions of the loxP site. The sequence entered into the search algorithm is shown below:

Pattern #1: ATAACTTCGTATA (n) {8} TATACGAAGTTAT (SEQ ID NO:4).

The (n) {8} allows the program to substitute any eight nucleotides in the region between the two thirteen-basepair inverted repeats and only look for similarity to the thirteen-basepair inverted repeats. Both strands were searched and no gaps or extensions were allowed.

When the search was conducted allowing for a maximum of eight mismatches, a large number of hits were obtained in the primate database. The total number of sequences searched was 73,825, representing 118,684,866 basepairs of sequence. The hits obtained from this search were then reviewed to identify likely pseudo-loxP candidates. Sequences having exact matches of at least four or five nucleotides immediately adjacent to the core on each side were given preference because mismatches more than five nucleotides away from the core on either side may be tolerated to some extent by Cre recombinase. A similar search was undertaken with the rodent database.

Search Pattern #2 made use of additional search criteria derived from structural studies of Cre. The crystal structure at 2.4 angstrom resolution of Cre recombinase complexed with loxP DNA reveals that contact is made between Cre and its target site at certain bases (Guo et al., 1997, Nature 389, 40–46). Footprinting with Fe-EDTA using Cre bound to the loxP site also reveals points of contact between Cre and bases in the loxP site (Hoess et al., 1990, J. Mol. Biol. 216, 873–882). These bases can be weighted more heavily to favor matching with the wild-type site. The search formula for determining a fit to these structural criteria was as follows for the 34-basepair lox site:

Pattern #2: ATnACnnCnTATA nnnTAnnn TATAnGnnGTnAT (SEQ ID NO:5).

Again, both strands were searched and no gaps or extensions were allowed. A search demanding four or fewer mismatches with the specified 16 basepairs yielded an extensive list of matches with the extant DNA sequences.

Searches were done in GenBank in the Primate, Rodent, Invertebrate, Plant, Fungus, and Bacteria databases. Some of the sites identified using these methods are shown in FIGS. 8A and 8B. The core sequences are shown in boldface type.

Example 2

In vitro Excision Assay of Pseudo-lox Sites in Bacteria and Human Cells

The following example demonstrates that the pseudo-recombination sequences of the invention are functional as sites for recombination of a nucleic acid sequence by a site-specific recombinase.

A negative control plasmid, pLCG1 (FIG. 1A), was created by inserting a 4.3-kb XbaI-BspHI fragment containing the lacZ gene, encoding β-galactosidase, driven by the CMV promoter (from pCMVSPORT-βgal, Gibco/BRL) into the EcoRV site of pLitmus29 (New England Biolabs, Beverly, Mass.) in the opposite orientation to the LacZα gene already present in the plasmid. This plasmid was then used as a base for the construction of other plasmids used in the excision assay. A very similar negative control plasmid, pL2β50, was used in some of the experiments in place of pLCG1. Briefly, annealed oligonucleotides containing the lox sites being tested and a marker restriction enzyme site were directionally cloned into the BamHI-HindIII sites on one side and the BglII-XbaI sites on the other side of the CMV-lacZ construct. This cloning was carried out to ensure that Cre-induced site-specific recombination would result in excision of the lacZ marker gene. A schematic representation

of the plasmids is shown in FIGS. 1A through 1C. FIG. 1D shows the DNA sequences of the lox sites from pWTLox[2] shown in FIG. 1B (top line of FIG. 1D) and plasmid p ψloxh7q21 shown in FIG. 1C (bottom lines of FIG. 1D).

The positive control plasmid used in the excision assay (pWTLox[2] FIG. 1B) had the 34-bp wild-type loxP site cloned into both the BamHI-HindIII site and the BglII-XhoI site. The test plasmids had a pseudo-recombination site cloned into the BglII-XhoI site and a recombination site containing the 13-bp palindromic repeats of loxP flanking the core sequence of the pseudo-recombination sequence cloned into the BamHI-HindIII site.

The bacterial strain used for the excision assay, 294-Cre (Buchholz, et al, Nucleic Acids Research 24:3318–3319, 1996) has been designed to constitutively express Cre recombinase at 37° C.

Approximately 1 ng of the DNA being tested was electrotransformed into the 294-Cre strain of *E. coli* using the Bio-Rad Gene Pulser (BioRad Laboratories, CA) at a field strength of 12.5 kV/cm, with a capacitance of 25 μF and resistance of 200Ω. Aliquots of the transformation mix were spread on plates containing ampicillin (100 μg/ml), methicillin (100 μg/ml), and X-gal (60 μg/ml). The plates were incubated at 37° C. for 18 hours, after which they were scored for the presence of blue and white colonies. Bacteria containing the parent plasmid pLCG1 generated a blue bacterial colony when grown on these plates, whereas bacteria containing a plasmid from which lacZ sequence has been excised generated a white colony. The excision frequency was defined as the ratio of the number of white colonies to the total number of colonies, expressed as a percentage.

As shown in Table 1 below, the excision frequency was close to 100% when the wild-type loxP sequences were present on the plasmid (positive control) and no excision was observed when no loxP sites were present.

TABLE 1

lox Site Tested	Mean Recombination Efficiency (%)
none	0.00
loxP	98.9
ψlox h7q21	11.5
ψlox h7q31	8.9
ψlox hXp22	99.0
ψlox h5p15	1.4
ψlox m9	4.0
ψlox m5	98.7

The results above are based on from 4 to 13 separate experiments for each plasmid tested. The data indicate that pseudo-recombination sequences are functional, and some pseudo-recombination sequences (ψlox hXp22 and ψlox m5) promote recombination at very high frequencies, comparable to the wild-type loxP sequence.

In conjunction with the data of Example 1, these recombination efficiency results help identify which basepairs within loxP are most critical for Cre binding. A strict correlation between the number of mismatches and the recombination efficiency was not observed. Therefore, it is clear that matches at specific positions are more important than overall homology. These results are consistent with the idea that the four bases flanking the core are important, as the ψlox h5p15 site, that has a mismatch in this region while otherwise having good matches, had the lowest recombination frequency. The wild-type core sequence was not required. For example, ψlox m5, which had a recombination frequency indistinguishable from that of loxP, had no matches to loxP in the 8-bp core. However, the best sites had only A and T basepairs in the central two positions of the core, indicating that this feature may be important.

The four ψlox sequences identified by using Pattern #2, ψlox hXp22, ψlox h5p15, ψlox m5, and ψlox m9, included the two ψlox sites with the highest excision efficiencies, ψlox hXp22 and ψlox m5, indistinguishable from loxP. On the other hand, ψlox h5p15, also obtained using Pattern #2, had the lowest recombination efficiency of the sites tested, probably because it contained a mismatch in the four positions nearest the core. These results suggest that while these first four positions are critical, the requirement for matching at the first five positions, used in screening the sites obtained with search Pattern #1, was overly restrictive. Good results would be obtained by using Pattern #2 in combination with a stringent requirement for matching at the first four positions from the core.

A similar assay was carried out in mammalian cells. Briefly, a plasmid expressing Cre, pBS185 (Life Technologies Inc., Grand Island, N.Y.) was modified by the insertion of a kanamycin resistance gene into the unique ScaI site to create pBS185-Kan. This modification renders cells transfected with plasmid resistant to kanamycin but sensitive to ampicillin. Approximately 2 μg of plasmid pBS185-Kan and 50 ng of one of the plasmids used in the bacterial assay described above were transfected into 293 (ATCC Accession No. 1573), human embryonic kidney cells, using LipofectAmine (Life Technologies) following the manufacturer's recommendations. The transfected cells were treated with DNaseI 24 hours after transfection. The cells were grown at 37° C. in Dulbecco's Modified Eagle medium (DMEM) for 72 hours after which low molecular weight DNA was isolated from the cells by Hirt extraction (Hirt, J. Mo. Biol. 26:365–369, 1967). The plasmid DNA was electrotransformed into *E. coli* strain DH10B (Life Technologies) under the conditions described above. Aliquots of the transformed bacteria were grown on amp/meth/X-gal plates as described above and scored for the presence of blue and white colonies.

Exemplary results are shown in FIG. 2. The frequency of excision seen in a mammalian cell background demonstrates the predictive nature of the bacterial assay system and demonstrates that the pseudo-recombination sequences of the invention are active substrates for recombinase-mediated recombination in a mammalian cell environment.

The ψlox h7q21 and ψlox hXp22 sites may mediate integration into the human genome. The ψlox h7q21 site is located in the q21 region of chromosome 7, while the ψlox hXp22 site is situated in band p22 of the X chromosome. The existence of these sequences in the human genome was verified by sequencing the appropriate PCR fragments covering the sites from human genomic DNA. Neither site is located in a coding sequence or a known gene.

Example 3

In vitro Transient Integration Assay of Pseudo-lox Sites in Human Cells

The following example provides a model system for assessing the ability of the pseudo-recombination sequences of the invention to promote genomic modification by site-specific insertion.

The ψlox site to be tested was placed on a plasmid having tetracycline resistance (FIG. 3, upper left). This plasmid

represented the chromosome and was the recipient for integration events. A lox site having the wild-type loxP palindromes and the 8-bp core of ψlox h7q21 was placed next to the lacZ gene on a second plasmid, this one having ampicillin resistance (FIG. 3, upper right). This plasmid represented the incoming donor vector. These plasmids were constructed as follows: The plasmid pTM1 was generated by cloning a 155 base-pair AflIII-SnaBI fragment from pLitmus29 containing the multiple cloning site into a unique EcoRV site of pUC-Tet, a tetracycline resistant derivative of pUC19 (C. R. Sclimenti and M.P.C., unpublished). The lox sites of interest were then cloned into the BglII-XhoI site of this plasmid to generate the recipient plasmids for the integration assay (pRWT and pRh7q21).

The plasmid pLGWTLox^2 was used as a base for the construction of the donor plasmids used in the integration assay. pLGWTLox^2 was created by treating pWTLox^2 with EcoRI and subsequent religation to excise the CMV promoter and create a unique EcoRI site between one of the loxP sites and the lacZ gene. Complementary oligonucleotides containing the loxP-derived palindromes with the core derived from the ψlox h7q21, a marker enzyme site, and EcoRI half-sites at the ends were annealed and ligated into the unique EcoRI site of pLGWTLox^2 to generate the pDh7q21 donor plasmid for the transient integration assay.

To perform the assay, 50 ng of the tetracycline-resistant recipient plasmid and 1 μg of the ampicillin-resistant donor plasmid were co-transfected into human 293 cells with Lipofectamine along with 2 μg of the Cre expression vector pBS185-Kan. The transfected cells were treated with DNaseI 24 hours after transfection. After 72 hours in human cells, plasmid DNA was purified by Hirt extraction (Hirt, J. Mo. Biol. 26:365–369, 1967) and returned to the DH10B strain of *E. coli* for detection of integration events. Plasmids that underwent integration were tetracycline resistant and now also carried lacZ (FIG. 3, lower left). They thus gave rise to blue colonies when plated on LB medium containing tetracycline and X-gal and incubated overnight at 37° C. Plasmid DNA was purified from blue colonies, and those plasmids with the restriction pattern expected for integration were classified as integrants. Each blue colony was restreaked on LB plates containing X-gal and either ampicillin and methicillin, or tetracycline. One representative plasmid was sequenced in the relevant regions to document integration at lox sites. The integration frequency was calculated as the number of integrants divided by the total number of tetracycline-resistant colonies.

The integration assay was performed with recipients bearing the ψlox h7q21 site or controls having either the wild-type loxP site or no lox site, along with the corresponding donors. The integration frequency at the wild-type loxP site was 0.41%. Integration at the ψlox h7q21 site was readily detectable and occurred at a frequency of 0.12%. Experiments performed with either the recipient alone or the donor alone in the presence or absence of the Cre expression plasmid did not yield any integrants. Transfection of the recipient and the donor in the absence of the Cre expression plasmid also failed to yield any integrants. These results demonstrate that detectable site-specific integration occurs at a pseudo-lox site in the human cell environment.

A second type of shuttle vector system that can be used to model chromosomal integration utilizes modified autonomously replicating vectors such as those described in issued U.S. Pat. No. 5,707,830. These types of vectors replicate stably in human cells and have a very low endogenous mutation frequency (DuBridge, et al, Mol. Cell. Biol. 7:379–387, 1987). Thus, they provide better models for the chromosome than newly transfected plasmid DNA. One preferred shuttle vector may have EBNA-1 sequences, the EBV family of repeats, oriP or a human chromosomal ori, a bacterial origin of replication, and a pseudo-lox sequence and a marker gene such as one conferring hygromycin resistance. This vector is established in mammalian cells using antibiotic selection. The cells are transfected with a plasmid expressing Cre and a plasmid having a lox recombination sequence and a second marker gene, such as a gene for chloramphenicol resistance. The assay is performed as described above.

Example 4

In vitro Chromosomal Assay for Integration Efficiency

The following example evaluates the efficiency at which a heterologous nucleic acid sequence can be inserted into a chromosome at a particular pseudo-recombination site (integration efficiency) and the level of expression of a gene sequence inserted therein.

Bicistronic assay vectors are constructed containing, for example, a gene coding for hygromycin resistance under the control of the thymidine kinase promoter and a gene encoding the enzyme chloramphenicol acetyl transferase (CAT) under the control of the cytomegalovirus immediate early promoter (Wohlgemuth, et al, Gene Therapy 3:503–512, 1996). The former marker is used primarily to assess integration frequency while the latter marker is useful for sensitively assaying the level and duration of gene expression. The vector additionally carries a lox sequence containing the core of the pseudo-loxP sequence under evaluation.

The test plasmid is transfected into mammalian cells, such as 293S cells (human) or NIH3T3 cells (mouse), along with a Cre-expressing plasmid, such as one of those described above. The transfected cells are grown in the presence of hygromycin and the number of hygromycin resistant colonies scored as a measure of integration frequency. A number of antibiotic resistant colonies are propagated and analyzed by polymerase chain reaction (PCR) and Southern blotting to determine whether they have an integration event targeted to the correct ψlox site. CAT gene expression is measured as follows. Cell extracts are prepared by standard procedures and total protein of the extract is normalized for total protein concentration and assayed for CAT activity as described by Gorman, et al, Proc Natl Acad Sci USA 79:6777, 1982 or Wohlgemuth, supra.

Example 5

In vivo Assay for Integration

The following assay evaluates the ability of a recombination sequence to promote integration of a heterologous nucleic acid sequence into a genome in vivo.

The in vivo integration and expression of the CAT gene by employing the teaching of the invention is evaluated essentially as described by Zhu, et al, Science 261:209–211, 1993. Vectors, one containing a lox recombination sequence and CAT gene and one expressing Cre, are mixed with liposomes that have a net cationic charge, for example, containing N[1-(2,3-dioleyloxyl)propyl]-N,N,N-trimethylammonium chloride (DOTMA) (Felgner, et al, Proc Natl Acad Sci USA 84:7413, 1987) and dioleoyl phosphatidylethanolamine (DOPE) in a 1:1 ratio. The ratio of DNA to liposomes is typically 1:1. The liposome/DNA mixture is typically injected into test mice in 200 μl of 5% dextrose in water intravenously through the tail vein.

At various time points, starting at 24 hours post-injection, test mice are sacrificed and various tissues harvested and homogenized. Cleared homogenates are assayed for CAT enzyme activity using a scintillation counting assay (Seed and Sheen, Gene 67:271–277, 1988) with the following modifications: 0.3 μCi of ^{14}C-labeled chloramphenicol (55 mCi/mmol) is added to 200 nmol of acetyl coenzyme A for a final volume of 122 μl. CAT activity is expressed as either CAT enzyme/weight of tissue or as a function of milligrams of protein in each tissue extract. Tissue extracts are prepared by standard procedures and total protein determined using standard protocols (Bradford, Lowrie, and the like).

Example 6

Intramolecular Integration Assay for a Site-Specific Recombinase in *E. coli*

The following example describes a rapid assay to measure site-specific integration by a recombinase. This assay was used to measure integration of the wild-type φC31 attB sequence into the wild-type φC31 attP sequence in the presence of the φC31 integrase. A similar assay can be used measure integration mediated by other recombinases of interest, such as the integrases of phages R4 and TP-901.

Integrase-expressing plasmids were constructed as follows. The φC31 integrase gene was amplified by the polymerase chain reaction from the plasmid pIJ8600 containing the φC31 integrase and attP (M. Bibb, John Innes Institute, Norwich, U.K.) with the following primers: 5'GAAC-TAGTCGTAGGGTCGCCGACATGACAC3' (SEQ ID NO:6) and 5'GTGGATCCGGGGTGTCTCGCTACGCCGC-TAC3' (SEQ ID NO:7). The PCR product was ligated into linear pCR2.1 (Invitrogen, Carlsbad, Calif.) at the T overhang to make the plasmid pTA-Int. The lacZ gene was removed from pCMVSPORTβGal (Life Technologies, Grand Island, N.Y.) by digestion with the restriction enzymes BamHI and SpeI, and replaced by the integrase gene from pTA-Int with BamH1 and Spe1 compatible ends, creating the plasmid, pCMVInt (FIG. 4B), which expresses φC31 integrase in mammalian cells under control of the cytomegalovirus immediate early promoter.

The integrase gene was subsequently removed from pCMVSPORTint by digestion with BamHI and PstI and ligated into pACYC 177 (resistances ampicillin and kanamycin) (S. Cohen, Stanford University, Stanford, Calif.) that had also been treated with BamHI and PstI, removing part of the ampicillin resistance gene. Finally, the lacZ promoter was removed from PBCSK+ (Stratagene, La Jolla, Calif.) by digestion with SacI and SapI. The integrase-containing pACYC plasmid was digested with PstI and SacI, and the lacZ promoter was inserted upstream of the integrase gene with a linker (5'GCTCGGCCAAAAAGGCCTGCA3' (SEQ ID NO:8), 5'GGCCTTTTTGGCCG3' (SEQ ID NO:9), creating the plasmid, pInt (FIG. 4A), expressing the φC31 integrase under control of the lacZ promoter.

The intramolecular integration assay plasmid was constructed as follows. The bacterial attachment site for φC31 (attB) was amplified by PCR from Streptomyces lividans genomic DNA (S. Cohen, Stanford University, Stanford, Calif.) with the primers: 5'CAGGTACCGTCGACGATG-TAGGTCACGGTC3' (SEQ ID NO:10) and 5'GTCGACATT-GCCCGCCGTGACCG3' (SEQ ID NO:11). This attB fragment was ligated into linear pCR2.1 at the T overhang sites to create the plasmid pTA-attB containing a 285 bp attB region. The phage attachment site (attP) was amplified by PCR from pIJ8600 with the primers 5'CGACTAGTACT-GACGGACACACCGAA3' (SEQ ID NO:12), 5'GTAC-TAGTCGCGCTCGCGCGACTGACG3' (SEQ ID NO:13) and ligated into linear pCR2.1 at the T overhang sites to create the plasmid pTA-attP, containing a 221 bp attP region. The lacZa was removed from pBCSK+ by digestion with PvuI and KpnI, treatment with T4 polymerase, and religation. The full length lacZ gene from pCMVSPORTβGal was removed by digestion with SpeI and HindIII and cloned into the SpeI and HindIII sites of the lacZa deficient PBCSK+ to make pBCβGal. The attP was then removed from pTA-attP by SpeI digestion and cloned into the SpeI site of pBCβGal. The attB was then removed from pTA-attB by SalI digestion and cloned into the SalI site of the attP containing pBCβGal, to create the assay plasmid PBCPB+ (FIG. 4C), in which the TTG cores of the att sites are in the same orientation. In addition, a control plasmid, PBCPB–, in which the att sites were in opposite orientations, was also constructed.

The pInt plasmid was then transformed into DH10B bacteria, grown under kanamycin selection, and made electrocompetent by a standard protocol. The resulting electrocompetent DHInt cells were used in the bacterial intramolecular integration assay, conducted as follows. 200 ng of the assay plasmid of choice was electroporated into DHInt cells, allowed to recover for one hour, spread on plates containing chloramphenicol and Xgal, and grown at 37° C. If an intramolecular integration event occurs, the lacZ gene located between the attB and attP sites will be excised, and a resulting colony will be white. The frequency of intramolecular integration was therefore calculated as the number of white colonies divided by the total number of colonies.

When this assay was carried out in DHInt bacteria using pBCPB+, all colonies were white, indicating efficient integration. Thousands of colonies were assayed for each plasmid tested. The same plasmid produced only blue colonies in DH10B bacteria, in the absence of the integrase gene. These results verify that the assay plasmid carried functional attB and attP sites and that the φC31 integrase functioned efficiently in *E. coli* with no added co-factors. In contrast, the plasmid pBCPB–, which carried the att sites in inverted orientation, resulted in blue colonies, because the lacZ gene was merely inverted, not excised, by the integration reaction. The assay plasmid with no att sites, pBCSK-βgal, also yielded only blue colonies in DHInt cells. Restriction enzyme digestion of plasmid DNA purified from a representative number of white colonies verified that the intramolecular integration reaction occurred as expected and resulted in deletion of lacZ between the attB and attP sites.

Example 7

Intramolecular Integration Assay in Mammalian Cells

The following example demonstrates the ability of phage φC31 integrase to integrate sequences site-specifically and efficiently in a mammalian cell environment.

To perform the intramolecular integration assay in human cells, the same PBCBP+ plasmid was used as in the bacterial assay of Example 6. The pCMVInt plasmid was substituted for pInt to ensure expression of φC31 integrase in mammalian cells. Subconfluent (60–80%) 60 mm plates of human 293 cells grown in DMEM supplemented with 9% fetal bovine serum and 1% penicillin/streptomycin were transfected with lipofectamine (Life Technologies) at a ratio of 6 μg lipofectamine per μg of DNA. Experiments were performed with 100 ng of the assay plasmid of interest and 2 μg of pCMVInt. Controls performed in each experiment

included no DNA, pCMVInt only, pBCSK-βgal (assay plasmid with no att sites), pBCSK-βgal+pCMVInt, and PBCPB+ alone.

Twenty-four hours after transfection, the medium was supplemented with 50 Units/ml of DNaseI to reduce the background of untransfected DNA. Three days after transfection, the cells were harvested and low molecular weight DNA was recovered by using the Hirt procedure (Hirt, J. Mo. Biol. 26:365–369, 1967). A portion of this DNA was electroporated into competent DH10B *E. coli* cells and spread on plates containing chloramphenicol and Xgal to select only for the assay plasmid. The intramolecular integration frequency was determined to be the number of white colonies divided by the total number of colonies.

Using this assay system in mammalian cells, the φC31 integrase was shown to catalyze recombination between the full-length attB and attP sites of PBCBP+ at a frequency of 50.6% (mean of 16 experiments, standard error=2.32%). This frequency is likely to be an underestimate as plasmid DNA that never came in contact with the φC31 integrase was probably present, despite efforts to remove untransfected DNA with DNaseI. It is clear that the φC31 integrase catalyzes efficient site-specific integration in mammalian cells.

To verify site-specific recombination, 96 white colonies were picked and plasmid DNA was prepared and examined by restriction digestion. Of these, 97% contained a plasmid that represented the expected site-specific recombinant. The remaining colonies contained plasmids that carried large rearrangements that disrupted lacZ. The low frequency rearrangement of transfected plasmids was observed with all plasmids, with and without integrase and att sites, and can be attributed to transfection-associated mutation of newly introduced DNA.

Example 8

Determination of the Minimal Sizes of Recombination Sequences

The following example describes the process for determining the minimal sequences needed for recognition and recombination by a site-specific recombinase. This process was used to determine the minimal wild-type attB and attP sequences functionally recognized by the φC31 integrase in bacterial and mammalian cell environments. A similar process can be used to identify the minimal sequences recognized by other recombinases of interest, such as the integrases of phages R4 and TP-901. The minimal attB and attP sequences can then be used to identify pseudo-recombination sequences, for example as described above for the Cre-lox system.

Prior to this study, the minimal sizes for the φC31 attachment sites, attB and attP, had not been determined. The attB site had been localized to approximately 280 basepairs and the attP region had been localized to 86 basepairs (Thorpe and Smith, Proc. Natl. Acad. Sci. USA, 1998). The intramolecular integration assay described in Example 6 was used to determine the minimal functional sizes for these att sites. Short double-stranded adaptor molecules containing att sites of various lengths were created by annealing single-stranded oligonucleotides. These shorter sites were used to replace the full-length att sites in the pBCPB+ assay plasmid, and recombination efficiencies were determined by electroporation into *E. coli*.

To determine the minimal function size of attB, the 278-basepair full-length attB surrounded by BamHI and HindIII sites was removed. This fragment was replaced by the series of synthetic shorter sites having ends permitting their orientation-appropriate cloning into pBCBP+. The resulting plasmids were electroporated into DHInt *E. coli* cells and recombinants were scored as white colonies, as described in Example 6 above. FIG. 5 (left side) shows the results of these experiments. AttB sites of 50, 40, 35, and 34 basepairs all provided full recombination function, i.e. they functioned at 100% of the efficiency of the full-length attB. Reduction of the site to 33 basepairs produced a marked decrease in recombination activity. Therefore, 34 basepairs was determined to be the minimal function size of attB.

Once attB was determined to be 34 basepairs long, attP was subjected to a similar set of reductions. The reduced attP sites were assayed on a plasmid carrying attB34 rather than full-length attB. To perform these experiments, the full-length attP surrounded by SacII and SpeI sites was replaced with a series of synthetic annealed oligonucleotides bearing ends permitting their correct orientation-specific cloning into pBCPB+-attB34. FIG. 5 (right side) depicts the results of these experiments. The function of attP dropped off as its size was reduced from 40 to 36 basepairs. The DNA sequence revealed that the 38 basepair site encompassed the major inverted repeat evident in attP. However, it was apparent from this data that the next two outermost basepairs conveyed some function (P39A&B). From this analysis, the minimal size of attP was determined to be 39 basepairs.

To determine the frequency at which the reduced att sites function in mammalian cells, the same panel of plasmids was analyzed by using the intramolecular integration assay described in Example 7. Each of the assay plasmids was transfected into human 293 cells along with pCMVInt. After 72 hours in the mammalian cells, the plasmid DNA was purified by the method of Hirt (Hirt, J. Mo. Biol. 26:365–369, 1967) and transformed into DH10B *E. coli* cells for scoring of recombinants. The results of these experiments showed that minimal sizes for attB and attP similar to those determined in *E. coli* also applied in mammalian cells. Approximately 60–90% of the efficiency of the full-length att sites was achieved with the same reduced att sequences that worked at 100% efficiency in *E. coli*, likely because the overall reaction is somewhat less efficient in the mammalian cell environment.

These experiments to determine the minimal sizes of attB and attP provided the information that these recombination sites had sizes of 34 and 39 basepairs, respectively. These sizes are similar to that of the 34-basepair loxP site. A recombination site of this size will possess active pseudo recombination sites in large genomes, such as those of mammals and most plants. Thus, it is statistically expected that the pseudo recombination sites for the φC31 integrase will occur in these genomes. These pseudo recombination sites represent targets for chromosome engineering.

Example 9

Determination of the Amount of Heterogeneity Tolerated in the Core Sequence of a Recombinase Site

The amount of heterogeneity tolerated in the 3-bp core sequence of the attB and attP sequences recognized by the φC31 integrase was determined. Similar methods can be used to determine the amount of core heterogeneity tolerated in the cores of other recombinases of interest, such as the integrases of phages R4 and TP-901.

The φC31 integrase catalyzes recombination between attB and attP sites. These sites have minimal functional lengths of

34 and 39 basepairs, respectively. While largely distinct in sequence, attB and attP share a three basepair common core sequence, TTG, that includes the crossover region. In the case of the 8-basepair core region of the loxP site targeted by Cre recombinase, it has been found that its sequence is largely unimportant, as long as it matches between the two recombining sites. To determine if this behavior applied to the core region of the attB and attP sites of the φC31 integrase, the effects of mutations within this core region were examined.

A panel of plasmids was generated in which either attB, attP, or both sites were altered with a specific single base change. These changes were then assayed with the intramolecular integration assay in *E. coli* described in Example 6. A recombination event results in excision of the lacZ gene located between the att sites. Thus, when an assay plasmid is transformed into bacteria expressing φC31 integrase, a site-specific recombination event is scored as a white colony.

The TTG core was mutated in each position individually to all other base possibilities. The effects of these mutations in attB were investigated when paired with a wild-type attP. Conversely, the effects of a mutant attP paired with a wild-type attB were measured. By combining attB and attP sites that contained identical mutations, it was determined whether the core region needed to only match to be effective in recombination.

To carry out these experiments, oligonucleotides bearing the mutations to be tested were synthesized in the context of attB34 or attP40 (see Example 8). The mutant oligonucleotides were annealed and cloned into the chloramphenicol-resistant intramolecular integration assay vector pBCBP+ to replace the wild-type attB or attP, as in Example 8. Individual plasmids containing the mutation of interest were assayed for recombination in *E. coli* strain DHInt, which carries the kanamycin-resistant integrase expression plasmid pInt, described in Example 6. Assay plasmid DNA (2 ng) was electroporated into DHInt, and after a 1 hour recovery period at 37° C. in rich media, the transformations were plated on LB agar containing 25 mg/ml chloramphenicol, 60 mg/ml kanamycin, and 50 mg/ml X-gal. The plates were incubated overnight (16–18 hours) at 37° C., after which blue and white colonies were counted. The recombination fraction was expressed as the percentage of white colonies out of total colonies. The results of these experiments are shown in FIG. 6.

The first and third positions of the core showed some flexibility, while the center position did not. The first position appeared to tolerate only pyrimidines; the CTG double mutant worked well. The third position of attP could be changed to any base, and to the other purine for attB. Overall, the pattern of base substitutions tolerated in the recognition sites for the φC31 integrase more closely resembled the degree of tolerance for substitutions typical of the outer palindromes, rather than the core, of the loxP site. Thus, unlike the situation in the Cre-loxP system, the φC31 integrase has strong base preferences within the cores of its attB and attP recombination sites, and merely matching any two three-basepair core sequences will not suffice to generate efficient recombination in this system.

Example 10

Bimolecular Integration Assay into a Model Chromosome in Mammalian Cells

The following example demonstrates the ability of phage φC31 integrase to integrate sequences site-specifically and efficiently into a model chromosome in a mammalian cell environment.

Example 7 demonstrated that the φC31 integrase efficiently catalyzed site-specific intramolecular integration in mammalian cells. The next step was to show that the integrase could catalyze efficient site-specific integration of exogenous DNA into mammalian chromosomes in cell culture. EBV-based plasmids provide easy and useful models for chromosomes. EBV vectors exist in the nucleus, replicate in synchrony with the chromosomes, and bear chromatin indistinguishable from that of the chromosomes. They can be easily purified from cells and transformed into *E. coli* for rapid scoring of integration events. Thus they have great utility in characterization of the integration reaction in human cells.

In these experiments, a kanamycin-resistant EBV plasmid was equipped with an attB site and established in human 293 cells to create a stable attB-containing human cell line. An ampicillin-resistant plasmid carrying attP and lacZ was then co-transfected into the attB cell line, along with a plasmid expressing the φC31 integrase. To assay for integration products, after three days plasmid DNA was extracted and transformed into bacteria. Blue colonies that grew on plates containing kanamycin, ampicillin, and Xgal were scored integrants, while total colony number could be obtained by plating on kanamycin alone.

The attB and attP plasmids needed for this study were constructed as follows. The target EBV based plasmids were based on p220.2 (DuBridge et al, 1987). The control plasmid p220K was made by inserting the kanamycin resistance gene from the Kan-resistant Genblock (Amersham Pharmacia, Piscataway, N.J.) into the XmnI site of the ampicillin resistance gene of p220.2. To make attB-containing p220 plasmids, the ampicillin-resistance gene of p220.2 was removed by digestion with BspHI. The kanamycin resistance gene described above was isolated by digestion with PstI, and cloned into amp-p220.2 with BspHI-PstI linkers (5'CATGAGGCCAAAAAGGCCTGCA3' (SEQ ID NO:14) and 5'GGCCTTTTTGGCCT3' (SEQ ID NO:15) to create the plasmid p220K. The full length attB was removed from the plasmid pTA-attB (Example 6) by SalI digestion and cloned into the SalI site of p220K, creating the plasmid p220KattBfull (FIG. 4D). The 35 base pair attB was cloned into the SalI and BamHI sites of p220K by using the oligonucleotides, 5' gatccgatatcgcgcccgggagccaagggcacgccatggcaccg 3' (SEQ ID NO:16) and 5'cgacggtgccacgggcgtgccctttggctcccgggcgcgatatcg3' (SEQ ID NO:17), creating the plasmid p220KattB35.

These EBV plasmids, p220K, p220KattBfull, and p220KattB35, were established in human 293 cells as follows. 293 cells were grown in DMEM containing 9% fetal bovine serum and 1% penicillin/streptomycin to ~70% confluency in a 100 mm plates. 8 μg of p220KattBfull, p220Kattb35, or the control p220K were introduced by transfection with lipofectamine according to the manufacturer's protocol. At 24 hours post-transfection, the cells were split 1:4, and at 48 hours post-transfection hygromycin selection (350 μg/ml) was begun. 11 to 14 days after starting selection the cells were expanded and frozen down.

The attP-containing plasmid pTSAD (FIG. 4E) was constructed as follows. A multiple cloning site (oligos: 5'AATTACCGCGGGGCGCGCCGTTTAAACGCATGCCAATTGGGCCGGCCG3' (SEQ ID NO:18) and 5'AATTCGGCCGGCCCAATTGGCATGCGTTTAAACGGCGCGCCCCGCGGT3' (SEQ ID NO:19) was cloned into the EcoRI site of the plasmid pWTLox^2 (Example 2) upstream of lacZ, regenerating one EcoR1 site. The attP site was removed from the plasmid pTAattP (Example 6) by digestion with EcoRI and cloned into the

regenerated EcoR1 site of pWTLox^2 to create the plasmid pES1. The lacZ promoter was removed from pBCSK+ by digestion with PvuII and SacII and cloned into pES1 which had been digested with PmeI and SacII. The region containing attP, the lacZ promoter, and the lacZ gene was removed by digestion with BamHI and BglII and cloned into the BamHI site of pTSA30 (Gregory Phillips, Iowa State University, Ames, Iowa) to create the donor plasmid pTSAD. pTSA30 and its pTSAD derivative are temperature sensitive for plasmid replication in *E. coli*.

To perform the integration assay, EBV plasmid-containing cells were grown to confluency in DMEM containing 9% fetal bovine serum, 1% penicillin/streptomycin, and 200 μg/ml hygromycin in 10 cm plates. These plates were split into eight 60 mm plates and grown in the above medium without hygromycin for 24–48 hours, until they were approximately 60–80% confluent. pCMVInt (Example 7, FIG. 4B) and pTSAD were transfected in equimolar amounts (10 μg total DNA) using 50 μl Superfect (Qiagen, Valencia, Calif.) according to the manufacturer's protocol. As controls, no DNA, 4 μg pCMVInt, or 6 μg pTSAD were cotransfected with salmon sperm DNA (to 10 μg). In addition, an equimolar amount of a plasmid encoding the green fluorescent protein (a derivative of pEGFP-c1, Clonetech, Palo Alto, Calif.) with salmon sperm DNA to 10 μg was transfected in parallel into the EBV plasmid-containing cells to monitor transfection efficiency.

2.5–3 hours after transfection, the Superfect was removed from the cells and replaced with serum-containing medium. Cells were fed with medium containing serum and 50 U/ml 24 hours after transfection and harvested 72 hours after transfection. Low molecular weight DNA was purified by Hirt extraction (Hirt, J. Mo. Biol. 26:365–369, 1967) and transformed into DH10B *E. coli* by electroporation. Also, 24 hours after transfection, transfection efficiency was measured by counting the green fluorescent protein-expressing cells relative to the total number of cells. The transfection efficiencies typically ranged from 6–18%. Because untransfected cells would have no opportunity to undergo integration but would still contribute EBV plasmids to the bacterial assay in the form of white colonies, the transfection efficiency was needed to obtain the correct the integration frequency.

In a typical experiment, 15 μl of a transformation was spread on each of three plates containing kanamycin, Xgal, and IPTG, while 150 μl of the same transformation was spread on each of three plates containing ampicillin, kanamycin, Xgal, and IPTG. The bacteria were grown overnight at 42° C. for approximately 16 h. The elevated temperature prevented replication of pTSAD, which has a temperature-sensitive plasmid origin of replication. Integrants were scored as the blue colonies on the plates containing both kanamycin and ampicillin. Integration frequency was calculated as the number of blue colonies on kanamycin and ampicillin plates divided by the total number of colonies on kanamycin platesx10 for each set of transfections. Raw numbers for integration frequency were divided by transfection efficiency to obtain accurate values for integration frequency.

FIG. 7 lists the integration frequencies obtained with each of the EBV plasmids and the negative controls. Each line of the figure represents a minimum of three separate transfections. For p220K, which lacks the attB site, a negligible frequency of blue colonies was detected. Upon analysis, these plasmids were not integrants, but rather homologous recombination events that occurred through common amp sequences on the two plasmids. For p220KattB35, carrying a minimally sized attB, a significant number of blue colonies were detected. When corrected for the transfection efficiency in these experiments, the integration frequency was 1.7%. For p220KattBfull, the integration frequency was even higher, at 7.5%. This increase presumably reflects a favorable sequence context for the full attB site compared to the reduced site. Controls in which pCMVInt, pTSAD, and each of the EBV plasmids, p220K, p220KattBfull, and p220KattB35 were co-transformed directly into *E. coli* yielded negligible numbers of blue colonies (0.002% or less). These controls confirmed that the high frequency integration events scored above occurred in human cells, not in *E. coli*.

The integration frequency into an attB site located on an EBV plasmid is impressively high and several orders of magnitude higher than the frequencies of random integration or homologous recombination, highlighting the utility of this invention. Furthermore, the integrants are site-specific, as indicated by restriction mapping of more than 160 of the blue colonies from the experiments with p220KattB35 and p220KattBfull. In addition, two integrants each, from the experiments with p220Katt35 and p220Kattfull, were analyzed at the DNA sequence level across the junctions of the integration site, confirming that exact site-specific integration occurred between attB and attP. FIG. 7 indicates that, as expected, the reaction requires the presence of both the integrase gene (pCMVInt) and the attP target site (pTSAD). Because EBV vectors are nuclear, chromatinized minichromosomes, the high integration frequency obtained in this system is predictive of the expected integration frequencies into att sites located on the chromosomes.

Example 11

Assay for Integration into the Chromosomes of Mammalian Cells

The following example describes methods used to demonstrate the ability of phage φC31 integrase to site-specifically integrate sequences into mammalian chromosomes.

Cell lines carrying the wild-type φC31 attB site are prepared by transfecting human 293 cells with Lipofectamine and a plasmid carrying the attLB sequence and the hygromycin resistance gene. The cells are grown in DMEM containing hygromycin and resistant colonies propagated to mass culture. Integration of the attB sequence is verified by Southern blot analysis using plasmid sequences as probes. These cell lines are then transfected with Lipofectamine and a plasmid containing the attP sequence and a neomycin/G418 resistance gene and a plasmid expressing the φC31 integrase gene under control of the CMV promoter. The G418 antibiotic is added to the DMEM growth medium approximately 48 hours after transfection. Selection is maintained for approximately ten days, after which the number of colonies is scored.

Higher numbers of neomycin resistant colonies are seen in cells co-transfected with the φC31 integrase-expressing plasmid than in cells that do not receive the integrase. Likewise, higher numbers of neomycin-resistant colonies are obtained in cells lines carrying attB compared to the parent 293 cell line lacking attB. These results suggest that the φC31 integrase enzyme can catalyze the integration of heterologous sequences into a mammalian genome, both at an integrated attB sequence and at endogenous pseudo-recombination sequences.

Similar experiments can be conducted using cell lines carrying an integrated attP hygromycin-resistant plasmid,

followed by transfection with a neomycin-resistant attB plasmid, to demonstrate integration into the integrated wild-type attP and attP pseudo-sites. Furthermore, similar experiments can be conducted in other cell types, such as those derived from other mammalian species or from plants, to test integration activity in these cellular backgrounds.

While the foregoing has been with reference to particular embodiments of the invention, it will be appreciated by those skilled in the art that changes in these embodiments may be made without departing from the principles and spirit of the invention, the scope of which is defined by the appended claims.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 41

<210> SEQ ID NO 1
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Bacteriophage P1

<400> SEQUENCE: 1

ataacttcgt atagcataca ttatacgaag ttat 34

<210> SEQ ID NO 2
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Saccharomyces cerevisiae

<400> SEQUENCE: 2

gaagttccta tacttctaga agaataggaa cttc 34

<210> SEQ ID NO 3
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: phage R4

<400> SEQUENCE: 3

gaagcagtgg ta 12

<210> SEQ ID NO 4
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: loxP search pattern
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)...(34)
<223> OTHER INFORMATION: n = A,T,C or G

<400> SEQUENCE: 4

ataacttcgt atannnnnnn ntatacgaag ttat 34

<210> SEQ ID NO 5
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: loxP search pattern
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)...(34)
<223> OTHER INFORMATION: n = A,T,C or G

<400> SEQUENCE: 5

atnacnncnt atannntann ntatangnng tnat 34

<210> SEQ ID NO 6
<211> LENGTH: 30
<212> TYPE: DNA

-continued

<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: primer

<400> SEQUENCE: 6

gaactagtcg tagggtcgcc gacatgacac 30

<210> SEQ ID NO 7
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: primer

<400> SEQUENCE: 7

gtggatccgg gtgtctcgct acgccgctac 30

<210> SEQ ID NO 8
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: linker

<400> SEQUENCE: 8

gctcggccaa aaaggcctgc a 21

<210> SEQ ID NO 9
<211> LENGTH: 14
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: linker

<400> SEQUENCE: 9

ggccttttttg gccg 14

<210> SEQ ID NO 10
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: primer

<400> SEQUENCE: 10

caggtaccgt cgacgatgta ggtcacggtc 30

<210> SEQ ID NO 11
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: primer

<400> SEQUENCE: 11

gtcgacatgc ccgccgtgac cg 22

<210> SEQ ID NO 12
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: primer

<400> SEQUENCE: 12

cgactagtac tgacggacac accgaa 26

-continued

<210> SEQ ID NO 13
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: primer

<400> SEQUENCE: 13

gtactagtag cgctgcgcgc actgacg 27

<210> SEQ ID NO 14
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: linker

<400> SEQUENCE: 14

catgaggcca aaaaggcctg ca 22

<210> SEQ ID NO 15
<211> LENGTH: 14
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: linker

<400> SEQUENCE: 15

ggccttttg gcct 14

<210> SEQ ID NO 16
<211> LENGTH: 46
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: p220KattB35 first oligonucleotide

<400> SEQUENCE: 16

gatccgatat cgcgcccggg gagcccaagg gcacgcctg gcaccg 46

<210> SEQ ID NO 17
<211> LENGTH: 46
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: p220KattB35 second oligonucleotide

<400> SEQUENCE: 17

tcgacggtgc cagggcgtgc cttgggctc cccgggcgcg atatcg 46

<210> SEQ ID NO 18
<211> LENGTH: 48
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: multiple cloning site first oligonucleotide

<400> SEQUENCE: 18

aattacgcg gggcgcgcg tttaaacgca tgccaattgg gcggcgcg 48

<210> SEQ ID NO 19
<211> LENGTH: 48
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

-continued

<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: multiple cloning site second oligonucleotide

<400> SEQUENCE: 19

aattcggccg gcccaattgg catgcgttta aacggcgcgc cccgcggt 48

<210> SEQ ID NO 20
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: wild-type loxP site

<400> SEQUENCE: 20

ataacttcgt ataatgtatg ctatacgaag ttat 34

<210> SEQ ID NO 21
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: "psi" loxh7q21

<400> SEQUENCE: 21

atacatacgt atatatgtat atatacatat atat 34

<210> SEQ ID NO 22
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:" psi" coreh7q21

<400> SEQUENCE: 22

ataacttcgt atatatgtat atatacgaag ttat 34

<210> SEQ ID NO 23
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 23

acaaccattt ataatatata atatatgatg ttat 34

<210> SEQ ID NO 24
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 24

atacatacgt atatatgtat atatacatat atat 34

<210> SEQ ID NO 25
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 25

atatacacgt atatatatat atatacgtat atat 34

-continued

<210> SEQ ID NO 26
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 26

caaacaaggt atatgcctgt atatacgaaa tggt 34

<210> SEQ ID NO 27
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 27

atatatacgt atatatacat atatacgtat atat 34

<210> SEQ ID NO 28
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 28

atatatacgt atatatacat atatacacat atat 34

<210> SEQ ID NO 29
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 29

ataaatatgt atatgtatat gtatacgtat ataa 34

<210> SEQ ID NO 30
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 30

atatatatgt atatgtatat gtatacgtat atat 34

<210> SEQ ID NO 31
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 31

atatatacgt atatacacat atatacgtat atac 34

<210> SEQ ID NO 32
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Mus sp.

<400> SEQUENCE: 32

atattgacat atattataaa gtataagtag ttat 34

<210> SEQ ID NO 33
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Mus sp.

<400> SEQUENCE: 33

gtaactgagt atatgcatat atatacgtat atat 34

-continued

<210> SEQ ID NO 34
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Mus sp.

<400> SEQUENCE: 34

ataacatatt atatttatat atatatatat ttaa 34

<210> SEQ ID NO 35
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Mus sp.

<400> SEQUENCE: 35

atatatatgt atatatatac atatacatac atat 34

<210> SEQ ID NO 36
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Mus sp.

<400> SEQUENCE: 36

agcacttcct atataacttc atatacgtag ctcc 34

<210> SEQ ID NO 37
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Caenorhabditis elegans

<400> SEQUENCE: 37

atagcgtcgt ataatccgaa atatacagat ctat 34

<210> SEQ ID NO 38
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Arabidopsis thaliana

<400> SEQUENCE: 38

ctagtttggt atatatatat atatactaat ttat 34

<210> SEQ ID NO 39
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Arabidopsis thaliana

<400> SEQUENCE: 39

ataactttgt atagtttaac ttatattagg tact 34

<210> SEQ ID NO 40
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Flaveria anomala

<400> SEQUENCE: 40

atcagttagt atatattcgt atatacgtag atat 34

<210> SEQ ID NO 41
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Saccharomyces cerevisiae

<400> SEQUENCE: 41

ttgccttcgt atataccttt ctataccaag taat 34

What is claimed is:

1. A method of site-specifically integrating a polynucleotide sequence of interest in the genome of an isolated eucaryotic cell, said method comprising:

introducing (i) a circular targeting construct, comprising a first recombination site and the polynucleotide sequence of interest, and (ii) an expression cassette comprising a polynucleotide encoding a site-specific recombinase into the isolated eucaryotic cell, wherein (a) the genome of said isolated eucaryotic cell comprises a second recombination site native to the genome, (b) recombination between the first and second recombination sites occurs in the presence of the site-specific recombinase, and (c) the site-specific recombinase is selected from the group consisting of φC31 phage recombinase, TP901-1 phage recombinase, and R4 phage recombinase; and

maintaining the isolated eucaryotic cell under conditions that allow recombination between said first and second recombination sites, wherein the recombination is mediated by the site-specific recombinase and the recombination results in site-specific integration of the polynucleotide sequence of interest in the genome of the isolated eucaryotic cell.

2. The method of claim 1, wherein said first and second recombination sites are a bacterial genomic recombination site (attB) and a phage genomic recombination site (attP).

3. The method of claim 2, wherein (i) said second recombination site comprises a pseudo-attP site, and (ii) said first recombination site comprises the attB site.

4. The method of claim 3, wherein said site-specific recombinase is selected from the group consisting of φC31 phage recombinase, TP901-1 phage recombinase, and R4 phage recombinase.

5. The method of claim 2, wherein (i) said second recombination site comprises a pseudo-attB site, and (ii) said first recombination site comprises the attP site.

6. The method of claim 5, wherein said site-specific recombinase is φC31 phage recombinase.

7. The method of claim 5, wherein said site-specific recombinase is R4 phage recombinase.

8. The method of claim 5, wherein said site-specific recombinase is TP901-1 phage recombinase.

9. The method of claim 2, wherein (i) attB comprises a first DNA sequence (attB5'), a bacterial core region, and a second DNA sequence (attB3') in the order attB5'-bacterial core region-attB3', (ii) attP comprises a first DNA sequence (attP5'), a phage core region, and a second DNA sequence (attP3') in the order attP5'-phage core region-attP3', and (iii) the recombinase mediates production of recombination-product sites that can no longer act as a substrate for the recombinase, said recombination-product sites comprising the order attB5'-(recombination-product site)-attP3' and attP5'-(recombination-product site)-attB3'.

10. The method of claim 9, wherein (i) said second recombination site is a pseudo-attP site, and said second recombination site comprises a first DNA sequence (attT5'), a core region B, and a second DNA sequence (attT3') in the order attT5'-core region B-attT3', (ii) said first recombination site is an attB site comprising attB5'-bacterial core region-attB3', in the order recited and (iii) the recombinase mediates production of recombination-product sites that can no longer act as a substrate for the recombinase, said recombination-product sites comprising the order attT5'-(recombination-product site)-attB3'{polynucleotide of interest}attB5'-(recombination-product site)-attT3'.

11. The method of claim 9, wherein (i) said second recombination site is a pseudo-attB site, and said second recombination site comprises a first DNA sequence (attT5'), a core region B, and a second DNA sequence (attT3') in the order attT5'-core region B-attT3', (ii) said first recombination site is an attP site comprising attP5'-phage core region-attP3', in the order recited and (iii) the recombinase mediates production of recombination-product sites that can no longer act as a substrate for the recombinase, said recombination-product sites comprising the order attT5'-(recombination-product site)-attP3'{polynucleotide of interest}attP5'-(recombination-product site)attT3'.

12. The method of claim 1, wherein said circular targeting construct further comprises a bacterial origin of replication.

13. The method of claim 1, wherein said circular targeting construct further comprises a selectable marker.

14. The method of claim 13, wherein said selectable marker provides for either positive or negative selection.

15. The method of claim 1, wherein said polynucleotide sequence of interest comprises a promoter sequence.

16. The method of claim 1, wherein said polynucleotide sequence of interest comprises at least one expression cassette.

17. The method of claim 16, wherein said expression cassette of said polynucleotide sequence of interest comprises a promoter operably linked to a polynucleotide sequence that encodes a product.

18. The method of claim 17, wherein said product is an RNA molecule.

19. The method of claim 17, wherein said product is a polypeptide.

20. The method of claim 1, wherein the expression cassette comprising a polynucleotide encoding the site-specific recombinase is carried on a transient expression vector.

21. The method of claim 1, wherein said expression cassette comprising a polynucleotide encoding the site-specific recombinase is introduced into the isolated eukaryotic cell before introducing the circular targeting construct.

22. The method of claim 1, wherein said expression cassette comprising a polynucleotide encoding the site-specific recombinase is introduced into the isolated eukaryotic cell concurrently with introducing the circular targeting construct.

23. The method of claim 1, wherein said expression cassette comprising a polynucleotide encoding the site-specific recombinase is introduced into the isolated eukaryotic cell after introducing the circular targeting construct.

24. A vector for site-specific integration of a polynucleotide sequence into the genome of an isolated eucaryotic cell, said vector comprising,

(i) a circular backbone vector,

(ii) a polynucleotide of interest operably linked to a eucaryotic promoter, and

(iii) a single recombination site, wherein said single recombination site comprises a polynucleotide sequence that recombines with a second recombination site in the genome of said isolated eukaryotic cell and said recombination occurs in the presence of a site-specific recombinase selected from the group consisting of φC31 phage recombinase, TP901-1 phage recombinase, and R4 phage recombinase.

25. The vector of claim 24, wherein said circular backbone vector is a procaryotic or eucaryotic vector.

26. The vector of claim 24, wherein said polynucleotide of interest operably linked to a eucaryotic promoter further comprises additional control elements.

27. The vector of claim 24, wherein the site-specific recombinase is φC31 phage recombinase.

28. The vector of claim 24, wherein said first and second recombination sites are a bacterial genomic recombination site (attB) and a phage genomic recombination site (attP).

29. The vector of claim 28, wherein said first recombination site is either attB or attP.

30. The vector of claim 29, wherein said recombinase is the site-specific ϕC31 phage recombinase.

31. The vector of claim 24, wherein said circular backbone vector further comprises a bacterial origin of replication.

32. The vector of claim 24, wherein said circular backbone vector further comprises a selectable marker.

33. The vector of claim 32, wherein said selectable marker provides for either positive or negative selection.

34. A kit for site-specific integration of a polynucleotide sequence into the genome of an isolated eucaryotic cell, said kit comprising,

(i) a vector of claim 24, and

(ii) a polynucleotide encoding a site-specific recombinase, wherein recombination between the first and second recombination sites occurs in the presence of the site-specific recombinase and said site-specific recombinase is selected from the group consisting of ϕC31 phage recombinase, TP901-1 phage recombinase, and R4 phage recombinase.

35. A method of modifying a genome of an isolated eucaryotic cell, said method comprising the steps of

(a) providing an isolated eucaryotic cell that does not comprise an attB or attP recombination site recognized by a site-specific recombinase selected from the group consisting of ϕC31 phage recombinase, TP901-1 phage recombinase, and R4 phage recombinase; and

(b) inserting an attB or an attP recombination site into the genome of the isolated eucaryotic cell, wherein said recombination site is recognized by said site specific recombinase, thereby modifying the genome of the eukaryotic cell.

36. The method of claim 35, wherein said inserting in step (b) is carried out by transforming the cell with a polynucleotide containing the attB or attP recombination site under conditions such that the polynucleotide is inserted into the genome.

37. The method of claim 35, further comprising

introducing (i) a circular targeting construct, comprising an attP recombination site and a polynucleotide sequence of interest, and (ii) an expression cassette comprising a polynucleotide encoding the site-specific recombinase into the isolated eucaryotic cell, of step (b) and recombination recombinase,

maintaining the isolated eukaryotic cell under conditions that allow recombination between said attP and attB recombination sites, wherein the recombination occurs in the presence of the site-specific recombinase and the result of the recombination is site-specific integration of the polynucleotide sequence of interest in the genome of the isolated eukaryotic cell.

38. The method of claim 35, further comprising

introducing (i) a circular targeting construct, comprising an attB recombination site and a polynucleotide sequence of interest, and (ii) an expression cassette comprising a polynucleotide encoding the site-specific recombinase into the isolated eucaryotic cell, of step (b),

maintaining the isolated eukaryotic cell under conditions that allow recombination between said attB and attP recombination sites, wherein the recombination occurs in the presence of the site-specific recombinase and the result of the recombination is site-specific integration of the polynucleotide sequence of interest in the genome of the isolated eukaryotic cell.

* * * * *